Moral Reasons Arbitrariness

Brad Seeman
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Part of the Philosophy Commons

Recommended Citation
https://ecommons.luc.edu/luc_diss/210

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
Copyright © 2009 Brad Seeman
LOYOLA UNIVERSITY CHICAGO

MORAL REASONS ARBITRARINESS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

PROGRAM IN PHILOSOPHY

BY

BRADLEY NELSON SEEMAN

CHICAGO, IL

DECEMBER 2009
ACKNOWLEDGEMENTS

As I take the last few steps of a long journey, one of the greatest pleasures is to recall
the kind encouragement and assistance of many people along the way. The first steps of the
journey were with Dr. Andrew Cutrofello, my first professor at Loyola and the one who
cleared a path for me to understand many difficult texts, beginning with Kant’s Critique of the
Power of Judgment. So much of what Andrew has done from the start—his cheerful
willingness to talk, trading a continuous stream of e-mails, shepherding me through
administrative thickets, chairing my Masters Paper and being a member of my dissertation
committee, and especially his constant, critical encouragement of my work—all this and
more has made this journey possible. Truly I am grateful to have run into Andrew my first
day at Loyola.

My second semester at Loyola I found more reasons to be grateful as I had the
chance to take Dr. Tom Carson’s History of Ethics course. Tom has taught me so much
about ethical theory, sometimes with exasperated shakes of his head and speechless
amazement at the angles I took toward things from off the beaten path (he took to calling
me J.L. one semester because of my fondness for Mackie’s queerness argument and my
determination to turn it around and run it toward theism). Tom has taught me to be more
careful in the ways I formulate my positions (a work in progress, I’m sure he’d say), which is
a helpful discipline for someone given to bigger pictures and trying to slash (sometimes too
quickly) to the heart of the matter. I have grown as a philosopher from Tom’s patient input.
Also formative for me was having Dr. Adrian Peperzak’s Heidegger course and Dr. Jennifer Lackey’s Philosophy of Mind course (during a semester at Northern Illinois University) in rapid succession. Peperzak is as “continental” as they come, and Lackey was unbendingly “analytic,” but both of them were excellent teachers, each with a different way of approaching their topics. And I thoroughly enjoyed the surprising connections that could be seen in juxtaposing these opposing ends of the philosophical spectrum, some of which shape this dissertation. They each also set models of excellent teaching that I have tried to emulate.

It was not until the end of my class work that I finally could make my schedule work to take a class with Dr. Paul Moser. I came to Loyola with the hopes of studying with Paul, but failed to make my teaching schedule work (and in my attempts to twist his arm for a reading course in early analytic philosophy… ah, well). When the opportunity came at long last, I was not disappointed. I always took the train home playing with all sorts of new and fruitful ideas about philosophy and about my Christian faith. Paul also provided a place of freedom to explore connections between theology and philosophy, being as convinced as I am that the connections between them are rich and fruitful. In his teaching and in his willingness to chair my dissertation, I have found more reasons to be thankful.

I also owe a great debt to friends who have encouraged me and prayed for me along the journey. These friends include Keith Crispin, Dr. Perry and Sandy Downs, Joel Gegner, Joe Kuleza, Steve Mobile, David Tsai, and Dr. John and Susan Woodbridge. My sister and brother-in-law, Susan Madsen and Rick Madsen also prayed faithfully for me and offered their encouragement. Also, I am thankful to be a part of a supportive and collegial department at Taylor University. They have encouraged me greatly. I have also benefitted
greatly from the efforts of Carol Warner and Jo Ann Cosgrove, the Interlibrary Loan Librarians at Judson University and Taylor University respectively.

The most important people in guiding me and encouraging me in the journey have been my mom, Sheila, and my dad, Russ. They prayed continually for me, financed my education, sacrificed promotions so as not to uproot us in the junior high years, took us to loving and supportive Churches where God was honored, calmly endured my rebellion and laziness, encouraged me to pray and read Scripture, listened and talked with me about everything under the sun, and just generally loved me. Dad even slogged through this entire dissertation when it was not a professional obligation (now that’s sacrificial love). I hope some of what I do with all that they have given me will bear fruit that will be to their credit and God’s glory.

In the last sixteen years, Kiersten has come along beside me and walked with me on the road. She is my wife, my companion, and my best friend. We have been blessed with five wonderful children, Grace, Emma, Luke, Elias, and Josiah. Together they have endured eight years where their husband or Daddy was always busy, often preoccupied, and sometimes crabby. Through it all they have been a source of great joy, good memories, and unflagging encouragement. Instead of sniping or griping, they have built me up and helped me continue moving forward. Kiersten in particular has loved me in many practical, tangible ways. I am grateful for her. “Her children arise and called her blessed; her husband also, and he praises her” (Proverbs 31:28).

Lastly, I think that it is of first importance in writing “acknowledgements” to give praise to the Lord God—Father, Son, and Holy Spirit. “Come to me, all you who are weary and burdened, and I will give you rest. Take my yoke upon you and learn from me, for I am
gentle and humble in heart, and you will find rest for your souls. For my yoke is easy and my burden is light,” (Matthew 11:28-30). Praise to you, Lord Christ.
For Russ and Sheila Seeman,

who have loved me and prayed for me for many years;

and for Kiersten,

who is the answer to so many of those prayers.
If we try to paint normative life as a part of nature, crucial parts keep looking off shape.

Reasons in the picture look not quite like genuine reasons…

Allan Gibbard

*Wise Choices, Apt Feelings*, p. 23
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................ iii

CHAPTER ONE: Introduction: Upon Encountering a Coal Pit .......... 1

CHAPTER TWO: Moral Reasons Arbitrariness ................. 23

I. What Is a Moral Reason? ........................................ 26
   A. Three Cases of Reasons-Talk .......................... 26
   B. Reasons and Causes .................................... 43
   C. Internalism and Externalism: Sketching the Topography of Reasons 62

II. Korsgaard and Harman, Dictators and Criminals: An Initial Approach to the 77
    Problem of Moral Reasons Arbitrariness
   A. Korsgaard’s Worry ....................................... 77
   B. Dismissing the Worry: Gilbert Harman’s Critique of the Justifying 81
      Reason/Motivating Reason Distinction
   C. Some Test Cases ........................................ 101

III. The Problem of Moral Reasons Arbitrariness .................. 106
   A. Two Kinds of “Arbitrariness” ............................ 106
   B. Deliberative Open-Endedness as an Unavoidable Feature of Human Morality 107
   C. Moral Reasons Arbitrariness ............................. 113

IV. Chapter Summary and a Look Ahead ........................... 132

CHAPTER THREE: Gibbard on Rationality, Plans, and the “Objective Pretensions” 136
   of Normativity

I. Rationality and Reasons in Gibbard’s Wise Choices, Apt Feelings 142
   A. Gibbard’s Norm Expressivism .......................... 142
   B. Gibbard on the “Objective Pretensions” of Morality 157
   C. Gibbard’s Problem of Circumscribing Acceptance and the Specter of Moral 165
      Reasons Arbitrariness
   D. Anorexics and Criminals, Dictators and Ruminators 188
   E. Slaveholders and Abolitionists: A Test Case for Gibbard’s Moral Philosophy 198

II. Thinking How To Live ....................................... 226
   A. Plans, Properties, and Concepts ........................ 226
   B. Settling the Telos of a Plan within a “Galilean” Metaphysic 233

III. External Reasons and the Second Person Address in Gibbard and Darwall 253
<table>
<thead>
<tr>
<th>CHAPTER FOUR: Korsgaard on Reasons, Reflective Endorsement, and the “Modern Scientific World View”</th>
<th>262</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Acting for Our Reasons: Human Reflection and the Limits of Endorsement</td>
<td>270</td>
</tr>
<tr>
<td>A. Value, Humanity, and the “Modern Scientific World View”</td>
<td>270</td>
</tr>
<tr>
<td>B. Identities and Maxims: Projecting the Logos</td>
<td>279</td>
</tr>
<tr>
<td>II. Identity Problems</td>
<td>298</td>
</tr>
<tr>
<td>A. The Identity/Reasons Problem</td>
<td>298</td>
</tr>
<tr>
<td>B. Setting Ends: Identity before Reasons</td>
<td>303</td>
</tr>
<tr>
<td>C. Setting Ends: Reasons before Identity</td>
<td>321</td>
</tr>
<tr>
<td>III. Korsgaard’s Kantian Transition Problem</td>
<td>331</td>
</tr>
<tr>
<td>IV. Korsgaard and Darwall</td>
<td>369</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER FIVE: From Millikan’s “Biological Norms” to Post’s Metaphysics of Morals</th>
<th>380</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Millikan’s Teleosemantics and the “Normativity Problem”</td>
<td>389</td>
</tr>
<tr>
<td>II. General Problems with Millikan’s System</td>
<td>413</td>
</tr>
<tr>
<td>A. Jackdaws, Bees, and Hormones; or, the “Quasi-” in Millikan’s “Quasi-Normativity”: Why We Should Keep It and the Problems It Causes for Millikan on Intentionality</td>
<td>414</td>
</tr>
<tr>
<td>B. Problems with Anchoring “Normativity” in History</td>
<td>431</td>
</tr>
<tr>
<td>C. How Is It That in Humans, Alone of All Animals, Imperatives Have Come Uncoupled from Indicatives?</td>
<td>434</td>
</tr>
<tr>
<td>III. Problems with Applying Millikan’s System to Morality</td>
<td>440</td>
</tr>
<tr>
<td>A. It’s a Long Way from Millikan’s “Normativity” to Moral Norms</td>
<td>440</td>
</tr>
<tr>
<td>B. “Pushmis,” “Pullyus,” and Morality</td>
<td>445</td>
</tr>
<tr>
<td>C. Is Moral Misrepresentation Possible?</td>
<td>449</td>
</tr>
<tr>
<td>D. “Meaning Rationalism” and Moral Reasons</td>
<td>452</td>
</tr>
<tr>
<td>IV. Post’s Attempt to Move from Nature to Norms</td>
<td>457</td>
</tr>
<tr>
<td>A. Paradox and “Pressing from Below”</td>
<td>461</td>
</tr>
<tr>
<td>B. The Adequacy of Reductions</td>
<td>468</td>
</tr>
<tr>
<td>C. Is There a “Double Standard”?</td>
<td>480</td>
</tr>
<tr>
<td>D. Justification and Authoritatively Prescriptive Moral Norms</td>
<td>485</td>
</tr>
<tr>
<td>E. When Owning Slaves Is Viable, Should One Be an Abolitionist?</td>
<td>499</td>
</tr>
<tr>
<td>V. Millikan, Post, and Moral Reasons</td>
<td>506</td>
</tr>
</tbody>
</table>
CHAPTER SIX: Conclusion: Being Led Out of a Coal Pit

I. Some Questions to Ask of Materialist Moral Philosophies

II. Two Concluding Questions
   A. How “Modest” Can Moral Objectivity Be before It Is Meager?
   B. How Might a Non-Materialist Moral Philosophy Address Moral Reasons Arbitrariness?

BIBLIOGRAPHY

VITA
CHAPTER ONE
INTRODUCTION: UPON ENCOUNTERING A COAL PIT

Thomas Reid once remarked about what he understood to be David Hume’s thoroughgoing skepticism that,

A traveller of good judgment may mistake his way, and be unawares led into a wrong track; and, while the road is fair before him, he may go on without suspicion and be followed by others; but, when it ends in a coal-pit, it requires no great judgment to know that he hath gone wrong, nor perhaps to find out what misled him.¹

I believe much moral philosophy today has led to a coal pit, and that we should turn around and seek a different path. The pit that concerns me has to do with a certain arbitrariness in the reasons for our moral actions. If we can avoid this arbitrariness, we should.

Now perhaps some philosophers would claim not to find the sort of moral arbitrariness that concerns me problematic; if so, they might greet what follows with a shrug and embrace the arbitrariness. In this Introduction, I argue that this is a very unpalatable stance. In order to do so, I need to give a brief statement of the kind of arbitrariness that concerns me, what I shall call “moral reasons arbitrariness.” I will not, however, be arguing or suggesting in this Introduction that any particular moral philosophy suffers from moral reasons arbitrariness or that there may not be any number of ways to avoid it (say, for example, full-information/ideal rationality theories). I mean only to help the reader see that if any moral

philosophy *were* to suffer from moral reasons arbitrariness, *that* moral philosophy would face a problem that cannot be lightly shrugged off.

If a moral philosophy suffers from moral reasons arbitrariness, it fails to establish support relations for moral judgments that uniquely justify those judgments in terms that make essential reference to a person’s ability to consider and weigh those support relations in making a decision about what should be done in a particular situation (as I will argue extensively in Chapter 2). This moral reasons arbitrariness can take different forms, including the direct recognition that one’s “moral reasons” for an action are rooted in factors adventitious to the consideration of support relations on the part of a person weighing what should be done in forming a moral judgment. Thus, when a person comes to see that the reasons she took to be in favor of torturing cats stem from nothing more than early formative experiences combined with a particular psychological disposition, she sees that what she had taken to be moral reasons are actually arbitrary, arising from forces adventitious to morality.

Moral reasons arbitrariness also includes coming to see that too much is “justified,” in the sense that other, conflicting moral judgments share the same justificatory grounds one’s own reasons are taken to enjoy. In this situation, you realize that the conflicting moral judgments of others are grounded in support relations that share the same form or type of provenance, differing only in non-moral contingencies that happened to play out

---

2 Chapter 2 will give careful attention to the notion of a reason (especially moral reasons), and will argue that a “support relation” is a key part of what a reason is.

3 In an insightful critique of the “Cornell realists” (or, as he calls them, the “synthetic realists”), Seiriol Morgan argues that their way of establishing “moral realism” inadvertently establishes a host of conflicting “realisms,” such as a Nietzschean “noble realism.” See Seiriol Morgan, “Naturalism and Normativity,” *Philosophy and Phenomenological Research* 72, no. 2 (March 2006): 319-44.
one way in your situation and another way in the other person’s situation. Their moral judgments enjoy as much justificatory ground as your own, and of a structurally identical kind, as when a person realizes that the consistency he took to ground his own abolitionism is shared by the slaveholder with whom he is arguing. And so, whether a moral philosophy inadvertently justifies too much and thereby undercuts the unique justificatory grounds of any particular moral judgment, or whether it directly exposes the justificatory inadequacy of what one had taken to be a moral reason, a moral philosophy that suffers from moral reasons arbitrariness fundamentally undermines the justificatory grounds of moral judgments. What had been taken to be justifying support relations for a moral judgment are “seen through,” so to speak. They are exposed as arbitrary.

Moral reasons arbitrariness results when a moral philosophy leaves those who adhere to it in a position where they see (or should see) that, according to their own moral philosophy, the support relations taken to be capable of addressing them in their deliberative capacity as they weigh what should be done morally do not uniquely support the moral judgments they are inclined to make. Non-moral forces intrude and moral judgments are revealed to be directly a result of such forces, or the support relations for one’s moral judgments are seen to arise from a justificatory procedure that also “justifies” other, conflicting moral judgments depending on how certain non-moral forces are arrayed, so that what one had taken to be moral reasons are highly suspect of being little more than detritus bobbing on the surface of deeper, non-moral forces—the same forces that carry others along as well as they make moral judgments. Once a moral philosophy allows moral “reasons” to become vehicles for the operation of non-moral forces that are themselves unresponsive to normativity, nothing obvious remains addressed to all people in a way that
appeals to or even commands (as opposed to bypassing) people precisely in their capacity to recognize, deliberate about, and respond to support relations precisely as supporting—again, even commanding—one moral judgment over another.

What drops out of such a moral philosophy is moral reasons of the right kind to address all people as moral agents and thus to provide a corrective (at least potentially) to the motivating reasons a person has due to the contents of his or her subjective motivational set (SMS). As a result, the grounds for reasoned moral critique go missing, with no independent ground for the correction of the vagaries of people's SMS's. In that case, whatever in fact motivates a person is a reason for that person. Of course from a moral standpoint the worry is that any such an SMS is down on all fours with any other SMS, leaving us in a situation where moral reasons are arbitrary. All such SMS's are determined by forces adventitious to the reason giving at the heart of our interaction as moral agents—unless there are reasons of the right kind that somehow stand apart from any particular SMS.

To put the same problem another way around, consider an example Allan Gibbard puts forward. In evaluating Hitler's reasons for invading Russia, I might understand his reason for doing so, but I wouldn’t say that this was a reason for doing so. Clearly Hitler's reason emerged from his SMS as a motivating reason and not a justifying reason. The key


6 A motivating reason is a reason a person herself finds for an action. Perhaps you stop to help someone change a flat tire because you like the way that person looks. This reason motivates you, but is does
question has to do with the status of my reasons for saying that his reason owed merely to his SMS, and was not in fact a justifying reason. Is my demurral rooted merely in my SMS, such that we have dueling SMS’s pitted against one another, with my preferences for the
“reasons” flowing out of my SMS owing simply to the (not very interesting) fact that they’re mine and I can’t avoid accepting them? In that case our moral reasons owe their force to contingencies alien to the moral realm that have shaped our SMS’s and are arbitrary precisely from the standpoint from which we take ourselves to be operating when we judge and act qua moral agents. But that makes our putative moral reasons “reasons of the wrong kind,” in Stephen Darwall’s helpful phrase, and there is no appeal outside of such reasons to an independent ground about which we may debate and argue in specifically moral and reasoned terms. This is moral reasons arbitrariness.

Now, it may be that when all is said and done some philosophers will see nothing else for it but to bite this bullet and admit that we are stuck with moral reasons arbitrariness. But bullets tend to explode when bitten, and this one packs a wallop. Moral reasons arbitrariness cuts the nerve of the human moral life by undermining the authority and obligation that is thought to attend moral judgments that are right. In a recent chapel service I heard the story of a girl in India who was drugged, locked up in the cellar of a brothel and beaten with a hose and drugged some more until she was eventually broken down into “servicing” twenty men a day. As it turned out, some people from International Justice Mission (www.ijm.org) were able to intervene and rescue this young woman. When

---

I hear of this, something in me cries out that her sexual slavery was wrong, and that any justification for such an action is simply mistaken and warped. Anyone who offers such a justification gets something wrong, where those who rescued the young woman have something right—and this owes to something much deeper than merely how things happen to stand with my SMS or the SMS of anyone else, no matter how widely shared or opposed. Simply put, no support relation exists that favors the sexual slavery of the girl, no matter who thinks otherwise. But in a situation of moral reasons arbitrariness, I am forced to see that—in point of fact—the moral reasons they offer for it being okay to enslave her stand on the same grounds as my own (the SMS they happen to have versus the SMS I happen to have), and it is difficult to see what would render one set of justificatory “moral reasons” more worthy of being chosen or acted upon that is not arbitrary precisely from a moral standpoint. In short, if moral reasons arbitrariness is not avoided, then in moral conflict it’s rational for the other person to respond to me in exactly the same way, with exactly the same grounds—and I know it, even if I can’t avoid taking my reaction as somehow superior.8

From my own perspective as a moral agent, the unique justification of my moral reasons is seen through and the prescriptive authority that is of the essence of being a moral reason dissipates, even where I cannot help but feel that my “reasons” must still be right. Indeed, from a standpoint within moral reasons arbitrariness, it becomes difficult to cash out what I would be violating if I were to act out of accord with my own best reasons, for I see that another set of moral reasons enjoys the same ground as those I currently accept. If I accept

---

8 In Chapter 3 we will see Gibbard making quite a deal about the fact that I cannot help but take my own reasons as being reasons for everyone, simply because that is part of what it is to accept them as reasons. The problem, of course, is that if I also believe that these “reasons” I accept are accepted simply because of the contingencies of my SMS, then I also have reason for seeing my own acceptance as morally arbitrary and thus not worth anything from a moral vantage point, since I see that ultimately it is a reason of the wrong kind.
moral reasons arbitrariness, I am forced to see that the authority and the obligation I take to
be characteristic of my moral judgments enjoys no unique grounds. After all, other,
conflicting judgments stand on the same grounds and have the same claim to “authority.”
My reasons, no less than yours, flow out of the vagaries of my SMS and have no standing
from within the moral standpoint we take ourselves to be standing within as we give moral
reasons to one another. Unless a moral philosophy provides moral reasons that somehow
stand over against our SMS’s and render some SMS’s well-formed, that moral philosophy
will be hard pressed to avoid moral reasons arbitrariness.

A moral philosophy plagued by moral reasons arbitrariness distorts a number of
important parts of human moral life. It will, for example, struggle to ground costly moral
action. A person who makes a moral judgment in a situation of moral reasons arbitrariness
will believe that her moral judgments arise directly from non-moral forces or that the
justificatory grounds she believes those judgments have also apply in the cases of people
who make conflicting moral judgments, the difference being in the way that non-moral
contingencies happen to have played out. It is readily understandable how this might make
a difference as she decides how much of a cost to pay for acting on those moral judgments.
After all, they she sees (or thinks she sees) that her moral judgments are not uniquely
justified, that others will look at her judgments and see the morally arbitrary grounds from
which they arise, or that they will be able to argue from structurally similar grounds for the
justification of their own conflicting views. Questions and doubts may arise: Are my moral
judgments really right? What if things had played out differently in my life… wouldn’t I
then have the same conviction for something else? Who—really—can rightly judge these
matters? Such nagging doubts gnaw at the grounds for costly moral action, and the Ouden
mallon of the ancient skeptic looms—“No more this than that”—attended by the “practical criterion” with its impulse toward the status quo. Moral reasons arbitrariness tends to undermine costly moral action, which is sometimes the hallmark of moral action that is importantly right and deserves moral praise.

Given all this, another deleterious effect of moral reasons arbitrariness on human moral life is that it’s not clear how moral “progress,” “blame,” and “praise” are to be understood. A moral philosophy marred by moral reasons arbitrariness makes it puzzling what moral progress I might make if I were to be persuaded by the moral reasons of someone else. Change, yes; progress, no. Nor does it make sense to praise a person for their “right” actions or to criticize them for their “wrong” actions, except as a compliment

---


In some cases, I would be happy to see this happen. I would like skinheads to find their views too costly and move toward the current status quo. On the other hand, there are some cases where I would think this would be a bad thing, such as with compassion the early Christians had on female babies left to die by exposure. Their action was costly—and right. Had they taken themselves to be in a situation of moral reasons arbitrariness, it seems doubtful that they would have been as inclined to pay the cost of their moral convictions. Movement toward the status quo at the hands of moral reasons arbitrariness would have been a moral loss.

10 From one angle one might be able to argue that the problem I am developing as “moral reasons arbitrariness” is a kind of skeptical “opposition” that, when a moral philosophy fails to address it adequately, ends up undermining moral judgments. However, in the case of some moral judgments, the argument ought to be run in the direction of *modus tolens* as exposing the weakness of the moral philosophies that cannot withstand the “opposition,” thus motivating a search for a more adequate moral philosophy. “If a moral philosophy is true it will vouchsafe external moral reasons that support moral judgments such as ‘Slavery is wrong’. This moral philosophy surrenders external moral reasons that support moral judgments such as ‘Slavery is wrong’ to moral reasons arbitrariness. Therefore, this moral philosophy is not true.”
paid to the likeminded or a hortatory device for the wavering or those easily cowed.\textsuperscript{11} No longer do I see such people as having gotten something right for which they are to be praised, or as having overlooked or turned away from a course of action that there were reasons for them to pursue; instead, they just happen to be in accord with me or not.

Criticism of moral error and admiration of moral praiseworthiness is localized and stripped of a unique ground in moral reasons. My “saint” is another’s “sinner,” and—if I accept moral reasons arbitrariness as being true—then I am forced to see that the only reasons I happen to see things this way are fully explicable in terms that are arbitrary with respect to moral reasons. Their reasons have the same ground as my own, and the reasons I offer for my actions fail to offer specifically moral grounds for those actions, but finally reduce to forces that are arbitrary when viewed from the standpoint of a moral agent. My best reasons are reasons of the wrong kind, with no apparent hope of finding anything else.

Likewise, and for the same basic reasons, people mired in a moral philosophy given to moral reasons arbitrariness will find it difficult to make sense of the ideas of moral error and moral discovery or recognition from within that moral philosophy (if, indeed, any philosophy were to suffer from the problem—as perhaps none do, or at least some may not). Does a murderer or a dictator fail to recognize reasons that she should have acted upon? What does she fail to recognize?\textsuperscript{12} What does she fail to discover that other people have found? If we are mired in moral reasons arbitrariness, it’s impossible to make sense of

\textsuperscript{11} In other words, we end up in the coal pit with Richard Rorty, having nothing more to do in our moral reason giving than to josh, cajole, bamboozle, seduce, and employ other means of coercion in getting others to see things our way.

\textsuperscript{12} Of course there are many candidates that have been put forward here, but that is beside the point in the present context, where we are presently considering only what follows for someone who accepts moral reasons arbitrariness, and not (yet) possible ways to avoid it. Of course these same problems would also accrue to moral philosophies that fail to avoid moral reasons arbitrariness.
these notions in the non-local way that morality demands. We may, of course, have
different moral concepts due to the different provinces in which our moral concepts were
formed, but these provincial moral reasons—seen to be such—are no longer thought to be
discovered or recognized; nor can those who do not share them be understood to be in
error in anything more than a provincial manner. Rather than being discovered or
recognized, our moral reasons just happen to be one way rather than another owing to the
ways our SMS’s happen to be constituted. They partake of no special status that does not
equally ground the moral reasons others give for contrary actions.13

In speaking of moral reasons being “recognized,” do we commit ourselves to a kind
of “infallibilism” about moral reasons? In one very basic sense, the answer is “no.” Our
awareness of moral reasons is often not clear, and all of us frequently make mistakes about
what we have reason to do in particular moral situations. When we think we are acting for a
moral reason (or the preponderance of such reasons), we are not necessarily correct about
that. Sometimes we think we have reasons for what we do, but we are incorrect.14 We are
only fallibly aware of our reasons (again, think of Hitler’s reasons for what he did as
opposed to what he had reason to do). But what is it to have a reason that we fail to

13 Perhaps a need to justify our SMS’s to others in a process of moral debate and criticism would rule
out many of the most objectionable SMS’s. If this were the case, then a moral philosophy that emphasized
this feature of moral discourse, perhaps by emphasizing ideal motives over actual motives, might not suffer
from moral reasons arbitrariness.

This introduction makes no claims about whether any moral philosophies actually do suffer from
moral reasons arbitrariness. Perhaps none do, for all that has been said to this point, and perhaps if some do
there are many different ways of avoiding the problem. The only point being made in this Introduction is that
if any moral philosophy were to suffer from moral reasons arbitrariness, that would count against it as a moral
philosophy. It would be better for a moral philosophy to avoid moral reasons arbitrariness. In part it is the
burden of this dissertation to argue that at least some materialist moral philosophies do in fact suffer from
moral reasons arbitrariness.

14 That this sounds both somewhat strange from one angle and perfectly normal from another points
to the fact that the notion of a “reason” is ambiguous, in this case between what are sometimes called
“motivating” and “justifying” reasons. This distinction and others will be explored in Chapter 2.
recognize? Think of a geometry problem (to take an example that will be developed in Chapter 2). As Jane works the problem, she may come to see that she has a reason to employ the Pythagorean Theorem—or she could fail to see that using the Pythagorean Theorem is what she should do. There is some sort of reason here that she can see or fail to see. We could speak of it “dawning” on her what she should do: “Oh, I see. I should use the Pythagorean Theorem right here.” Something about the reasons here is independent of Jane’s thought and stands over against her SMS. Whether or not Jane is motivated to employ the Pythagorean Theorem, there are still reasons to do so. Something is there to be recognized or discovered—something about which Jane can be right or wrong when she claims to have reasons for a particular step in working through the problem. As I will put it, the reasons are in some sense external to Jane.

There is, I want to say, something similar in moral cases (though it possesses a crucial additional dimension not found in the geometry example). This is explored in Chapter 2, where it will be argued that a moral philosophy runs a grave risk of moral reasons arbitrariness if it supports no such independence, no such externality to moral reasons. If a moral philosophy cannot sustain in the moral realm the more robust sense of reasons in view in the geometry example, then an important sort of constraint goes missing. Now of course a moral philosophy that results in moral reasons arbitrariness might be true, for all of this. But clearly there is not much to be said for morality if it is, for in that case the unique grounding of one course of action over another in the right kind of reasons cannot be sustained. The authority of moral reasons dissipates across a wasteland of conflicting reasons emerging from a welter of different SMS’s. Robust attributions of moral error lose the force of being uniquely grounded by moral reasons, as do attributions of
moral praiseworthiness. That which makes sense of such attributions—namely that genuine moral reasons have been discovered and acted upon or else mistaken or suppressed—has gone missing. The nerve of the human moral life is cut because it is seen not to be a matter of responding to the give and take of moral reasons. In light of all this, if we can find a view that allows us to avoid moral reasons arbitrariness rather than reworking the moral life in very different terms or dismissing it altogether, such a view has something very significant in its favor: it makes sense of a central part of the experience of at least many human beings. Moral reasons arbitrariness is to be avoided if at all possible.

In the pages that follow, I will defend the following thesis. Moral reasons arbitrariness is a serious problem for moral philosophy, one that vitiates the moral philosophies advanced by Allan Gibbard, Christine Korsgaard, and John Post (building on Ruth Garrett Millikan’s work).¹⁵ Their moral philosophies fail to avoid moral reasons arbitrariness because the ways they attempt to ground moral reasons cannot sustain reasons of the right kind that are external to us in the way requisite if there are to be reasons that stand over against the motivating reasons that owe to the contingent contents of our SMS’s. In particular, the moral philosophies of Gibbard, Korsgaard, and Post each succumb to moral reasons arbitrariness because they fail to underwrite independent, authoritatively prescriptive moral reasons. For purposes of this dissertation, then, I seek to establish (1) that moral reasons arbitrariness is a real problem, and (2) that Gibbard, Korsgaard, and Post

¹⁵ Gibbard and Korsgaard are widely recognized as important moral philosophers, and the reasons for careful attention to their work need no explanation. John Post is less widely known in this regard, but he is the only moral philosopher I know of to attempt to work the thought of one of the most important materialist philosophers working on the normativity of language and intentionality, namely, Ruth Garrett Millikan, into the foundations of a moral philosophy. As such his moral philosophy also merits attention, if for no other reason than to display some of the difficulties facing the attempt to press Millikan into service of moral philosophy. It might also display some challenges for a Millikan-type materialist in the realm of moral philosophy.
(Millikan) succumb to this problem due to flaws in the different ways they attempt to press rationality into the role of providing moral reasons. I believe that other interesting arguments lurk in the area, including (3) that external moral reasons\textsuperscript{16} of a specific sort are needed to avoid the problem, (4) that a certain kind of theism provides such reasons, and (5) that only theism can underwrite such reasons and that therefore only theism allows us to avoid moral reasons arbitrariness, but an attempt to argue that would obviously be too much for a dissertation. I do, however, intend to point toward what I believe to be a stiff challenge to moral philosophers (though it is not the business of this dissertation to argue for this challenge): in order to avoid moral reasons arbitrariness, a moral philosopher must vouchsafe external, authoritatively prescriptive moral reasons, and rationality, at least as employed in ways relevantly similar to the ways Gibbard, Korsgaard, and Post (Millikan) employ it, does not provide such reasons. Close attention to the struggles of Gibbard, Korsgaard, and Post (Millikan) goes some way toward making clear a problem I believe infects materialist\textsuperscript{17} moral philosophy generally: the inability to ground authoritatively prescriptive external moral reasons and a consequent inability to avoid moral reasons.

\begin{footnote}
\textsuperscript{16} Much will be done below to unpack the notion of an external moral reason; roughly, however, the idea will be that of an independent, authoritative prescriptivity.
\end{footnote}

\begin{footnote}
\textsuperscript{17} Materialism, as I am calling it here, is not an easy notion to pin down, as Barbara Montero and a number of others have made clear (see Barbara Montero, “The Body Problem,” \textit{Nous} 33 (1999): 183-200), and the ways in which the terminology is used invites confusion. Throughout the pages that follow, I will use the term “materialism” instead of “naturalism” to refer to the metaphysical stance that holds that non-personal entities or forces of a kind amenable to the current best physics or a future ideal physics are ontologically and temporally prior to persons and other living beings. I do this primarily to avoid any confusion with a quite different meaning of “naturalism” within ethical discourse, in which a “naturalist” position can very well refer in an essential way to God.

Some materialists don’t care for the term and prefer to call themselves “naturalists.” The chief reason offered seems to be that materialism has connotations of bulky, extended stuff—a notion fitting for 17\textsuperscript{th} and 18\textsuperscript{th} century materialists, but not for materialists living in the wake of Einstein, Bohr, and Heisenberg. Perhaps a certain palatability also factors into the choice, as “naturalist” is undoubtedly more winsome than “materialist.” In any case, any reader who wishes to substitute the word “naturalism” where I use the word “materialism” should feel free to do so—and with no need to sort out the metaphysical commitment (materialism) from the ethical position (naturalism).
\end{footnote}
arbitrariness. Again, I will not argue for this larger point here, but if I can carry out the more limited project of this dissertation in its entirety, or even in some substantial part, it will make substantial progress toward revealing what may be a much larger problem and will (in any case) be an original and important contribution to moral philosophy.

In trying to argue for my thesis, I will employ a fairly straightforward methodology. I will begin by clarifying key terms, placing my argument within the contemporary philosophical discussion, and showing how these important concepts and the ways in which the contemporary philosophical discussion has taken shape point to a serious problem: the problem of moral reasons arbitrariness. From there I will proceed to examine three materialist approaches to moral philosophy, those of Allan Gibbard, Christine Korsgaard, and John Post (who relies on Ruth Garrett Millikan) and expose deficiencies in the way they conceive of moral reasons by a close reading of their work. Each of these approaches gives

---

18 It should be noted (though I don’t like the terminology at all) that one need not be a “moral realist” in the common philosophical use of that terminology in order to ground the sort of authoritatively prescriptive moral reasons in view here. Moral realism is defined in a variety of ways, but I take the core notion as it has come to be formulated among philosophers to be the idea that moral judgments are true or false insofar as they are rightly related to facts about what is right or wrong that exist independently of the will or prescriptive address of any person whatsoever. Moral realism thus emphasizes the person-independence of what renders moral judgments right or wrong. This means that, according to today’s philosophical argot, there are theistic moral “non-realist” philosophies (for example, divine command theory) that would ground the kind of reasons needed to avoid moral reasons arbitrariness (if they could survive other challenges)—indeed, that might do it more successfully than competing views. I find calling such views “non-realist” to be misleading in the extreme. Tom Carson draws attention to this consequence of “moral realism” and points out that if we change the definition of moral realism so that it refers to “any human person” (or, as I would prefer, to any person who is not God), the modified definition would count such theories as versions of moral realism (see Thomas L. Carson, Value and the Good Life (Notre Dame, IN: Notre Dame University Press, 2000), 185). By Russ Shafer-Landau’s definition, then, I would be happy to call myself a moral realist (as I, indeed, find it natural to do): “Moral realism is the theory that moral judgments enjoy a special sort of objectivity: such judgments, when true, are so independently of what any human being, anywhere, in any circumstance whatever, thinks of them,” (Russ Shafer-Landau, Moral Realism: A Defense (Oxford: Clarendon Press, 2003), 2). I would accept only some such modified definition of moral realism, but, given the potential confusion surrounding the term “moral realism” I will avoid this terminology altogether.

19 For example, thought-provoking criticism of the important work of Gibbard and/or Korsgaard would be an important contribution to the field, and showing that Millikan’s work does not hold promise in moral philosophy by criticizing Post attempt to work her thought into a “metaphysics of morals” would also be helpful.
an inadequate account of moral reasons and ends up becoming mired in moral reasons arbitrariness. If all goes well, at the end we will be left with a clear challenge: either (1) find a materialist way to avoid the difficulties marring the account of moral reasons in the philosophies of Gibbard, Korsgaard, and Post/Millikan and their attempts to employ rationality to solve the problem of moral reasons arbitrariness, (2) collapse into moral reasons arbitrariness and start trying to spruce up the coal pit since these are our new digs, or (3) explore the possibilities for a non-materialist approach to moral philosophy, including an examination of a vibrant theistic moral philosophy that a number of other capable philosophers are exploring.

Employing this methodology, the argument will unfold in the following four chapters. The next chapter (Chapter 2) begins by looking at what reasons are and argues that authoritative prescriptivity should have a central place in our understanding of moral reasons. From there the chapter turns to consider something that worries Christine Korsgaard as she opens her argument in *The Sources of Normativity*. Essentially, Korsgaard has a gnawing suspicion that the “Why be moral?” question, taken as a question about moral justification, may have no good answer. She is, I think, right to be worried, and I will defend the importance of a distinction related to the one she is trying to draw—a distinction between justifying reasons and motivating reasons—from an attack by Gilbert Harman (though it will turn out in any case that Harman’s own position *illustrates* moral reasons arbitrariness and why it should be avoided). The important point here is that motivating reasons are not the only sort of reasons there are, and that a different kind of reasons is to be distinguished from motivating reasons. That is, as Stephen Darwall puts the matter, in moral discussion we refer “to *normative reasons* rather than *motivating reasons*, that
is, to reasons \textit{to} do something, rather than whatever reasons someone actually acts on or any motivational state that explains one’s act causally or teleologically.\textsuperscript{20}  Following this, I will work to untangle some of the knotted verbiage of contemporary debates about moral internalism and externalism, specifying a couple of issues (judgment internalism and reasons externalism) that I take to be of critical importance and stating how my argument will relate to them. The chapter draws to a close by showing how all of this leads to a very real problem of moral reasons arbitrariness and giving a brief survey that shows how widespread the problem is in contemporary moral philosophy.

The third chapter attempts to show how Allan Gibbard fails to avoid the problem of moral reasons arbitrariness, with help from two unsavory characters introduced in the second chapter: the contented criminal and the viable dictator. The first two sections of this chapter will examine the two main books in which Gibbard develops his moral philosophy. First, I will argue that in \textit{Wise Choices, Apt Feelings}, Gibbard’s notion of rationality—whatever its other merits—clearly fails to provide an adequate answer to the problem of moral reasons arbitrariness. Where Gibbard had hoped to salvage something substantive from what he calls the “objective pretensions” of normative judgment, he does not give a satisfactory account of the \textit{independence}, the \textit{interpersonal authority}, or the \textit{non-arbitrariness} that characterizes normative judgments—with the result that his moral philosophy suffers from moral reasons arbitrariness. This failure will be illustrated by an extensively developed example where I argue that Gibbard’s norm-expressivism cannot

\begin{footnotesize}
\begin{itemize}
\item[\textsuperscript{20}] Darwall, \textit{The Second-Person Standpoint}, 4, note 3; emphasis in original. I wonder, however, if Darwall goes a step too far in including the word “teleologically” here. This is certainly in line with his Kantianism, but I wonder if a teleologically grounded reason has to be a reason of the wrong kind. Or, to ask the same question in another way, I wonder if all teleological reasons could be reduced simply to motivating reasons. In any case, the distinction remains: motivating reasons do not swallow up the entire realm of reasons, Darwall wants to argue. In this he is definitely on solid ground, or so I will attempt to show in Chapter 2.
\end{itemize}
\end{footnotesize}
sustain a satisfactory objectivity to a moral judgment where it should: namely, that “slavery is wrong.”

The second section of Chapter 3 argues that Gibbard does no better in his recent book, *Thinking How to Live*. In both cases, Gibbard fails to show how human rationality as he understands it from within his materialist metaphysic can ever give us reasons that pack a normative punch. Reasons, as we see him worrying in the epigraph to this dissertation, look strangely misshapen and unable to function in a normative role—a role of justifying a course of action. The third section of this chapter compares Gibbard’s notion of the role of the second-person (the “you”) in moral reasoning with Stephen Darwall’s work on this matter. I argue that Darwall gets the better of this comparison, and that the reason why his position is more satisfactory is instructive for the problem I am pursuing in the dissertation. Darwall’s notion of second-personal reasons points in a helpful direction for finding reasons that are both external and of the right kind to address us *qua* moral agents, even though Darwall finally betrays his own best insights (as will be argued in Chapter 4). The conclusion of the chapter will be that in turning *outward* to the biological world of materialism what Gibbard finds there fails to ground moral reasons in a way sufficiently robust to avoid moral reasons arbitrariness. In short, Gibbard fails to transmute external natural *causes* he finds when he looks outward to the materialist world into external moral *reasons*.

Chapter 4 will consider Christine Korsgaard’s Kantian constructivism and her attempt to ground moral reasons by turning *inward* to the reason of the agent herself as that
fits into the natural world, particularly as this is tied to the maxims of free human thought. Korsgaard believes that internal moral reasons are quite robust, and that these need no foundation apart from the free thought of genuinely human beings to cut across the varying human motivational sets. Internal reasons capture all that external reasons supposedly offer in terms of criticism of moral error, making sense of moral reasoning, and the other desiderata of a moral theory that squares with human moral experience—and this without the metaphysical baggage of a notion of external moral reasons.

Much of Chapter 4 will be concerned to put pressure on Korsgaard’s notion of moral reasons by identifying a tension in her conception of maxims. It will be argued that Korsgaard wavers between what I call “identity-priority maxims” and “form-priority maxims,” with the tension in either conception pushing her toward the other. Additionally, in placing the bulk of the weight of her theory on the internal reasons of human beings formulated in the maxim-structure of thought, Korsgaard needs to show that the empirical realities of an evolved human psychology shaped by semantic-interpretational and cultural forces—as well as the vagaries of individual genetics and their bearing on the way different

21 The “inward turn”/“outward turn” language that will be employed at times in what follows will not be argued to present an exhaustive option, so that the only two possibilities for the materialist moral philosopher would be either to turn outward to the external, materialist/Darwinian world, or inward to the nature of the agent materialistically understood. I do in fact think that in some broad sense, and with many permutations of both possibilities and hybrids of the two, these are in fact the only possibilities open. Again, however, I will not be arguing that here, nor will I be relying on this position in anything that follows.

If the two options are exhaustive, an implication would be that if materialism creates a problem for each possible place the moral philosopher could turn to ground moral reasons, then materialism would be unable to avoid moral reasons arbitrariness. I do think that the arguments I make to show that moral reasons arbitrariness affects the moral philosophies of Gibbard, Korsgaard, and Post point suggestively to deeper problems materialists generally face whether they turn inward or outward to ground moral reasons. But I make no attempt to argue this in the pages that follow, nor is it in any way the burden of this dissertation to do so.

Thus the “inward turn”/“outward turn” language employed from time to time is used only as a description of what I take the individual philosophers below to be doing. For example, Korsgaard looks “inward” to the formal structure of the agent’s own maxims to ground moral reasons, rather than looking “outward” to ground moral reasons in facts about human evolutionary history, as Post does.
people reason—are not so protean that these internal reasons collapse under the weight she needs to place on them if she is to avoid the problem of moral reasons arbitrariness. In this connection it will be helpful to consider the fascinating possibility G.A. Cohen raises of an “idealized Mafioso” (the “contented criminal” of Chapter 2) whose moral reasoning is shaped by cultural forces that lead him to find no reason not to murder someone else. His internal reasons simply move in a different direction. Faced with this problem, Korsgaard can either retreat toward Kant and a more robust notion of reason, allowing her to hold that the Mafioso has external reasons to be moral (with these somehow attached to a Reason with no Reasoner), or else she can bite the bullet and allow that the Mafioso has no reasons to be moral. Korsgaard bites the bullet. And this is to succumb to the problem of moral reasons arbitrariness, I will argue.

As with the previous chapter, this chapter will conclude with Darwall’s work on second person reasons as a point of comparison. Again, I conclude that Darwall has important correctives we should heed, but this time I will argue that he and Korsgaard both finally succumb to the same difficulty and thus fail to avoid the problem of moral reasons arbitrariness. The conclusion of all this will be that Korsgaard’s inward turn to the formal features of a materialistically acceptable “moral standpoint” leaves moral reasons relative to the individual’s SMS. Korsgaard’s internal moral reasons are relative to individuals in a way external moral reasons would not be, thus leaving Korsgaard with a problem of moral reasons arbitrariness.

The fifth chapter returns to the attempt to look outward to the biological world of materialism to find a way to underwrite moral reasons (as Gibbard did), this time as it has been pursued by John Post, drawing on Ruth Garrett Millikan’s work on a materialistically
acceptable theory of intentionality. The essential ideas in this project are really Millikan’s, especially the notion of a “biological normativity” that emerges out of natural selection as that is considered diachronically, in its historical dimension. Millikan’s work is extremely technical and complex, and the first section of chapter five will be dedicated to understanding the broad outlines of her work, especially her insistence that reasoning is not done “in the head,” but in the world. Millikan is radically externalist. The second and third sections explore difficulties attending Millikan’s work, both in terms of her own project of formulating an acceptable materialist theory of intentionality and in terms of any attempt to apply her notion of “biological normativity” to the problem of moral reasons. Millikan herself does not make such an attempt, but John Post has recently pursued this angle in several articles and in his recent book, From Nature to Norm: An Essay in the Metaphysics of Morals. The huge difficulty facing Post (and Millikan) is how to transmute the causes of the biological world given to us by Darwin into the authoritative prescriptivity characteristic of moral reasons that address persons. Have Post and Millikan offered something that deserves to be called normativity? Chapter five will argue that the answer here is pretty clearly “No.” As they turn outward to the efficient causality of Darwin’s biological world, their various machinations leave them only with more causes—with the result that moral “reasons” are determined by adventitious causal forces. If reasoning is in the world and not in the head, as Millikan claims, then reasons will be whatever that biological world comes up with; moral reasons will be arbitrary from the perspective of the reasoning agent herself.

The dissertation concludes by reviewing the main lines of the argument, and then suggestively outlining a deeper challenge for materialist moral philosophers I believe my argument points to. If my arguments against Gibbard, Korsgaard, and Post/Millikan are
sound, then we have examples of three materialist philosophers who succumb to moral reasons arbitrariness. Furthermore, it is noteworthy that their capitulation to moral reasons arbitrariness comes as they follow two fundamentally different paths. Gibbard and Post/Millikan turn outward to the biological world of materialism and find only causes there—causes that they fail to show could underwrite the authoritative prescriptivity of moral reasons. Korsgaard pursues a different tack, turning inward to the formal features of our obvious ability to take what she calls the moral “standpoint”—a standpoint that sits very uneasily (if at all) in the materialist world to begin with, and that leaves moral reasons internal to individual persons in ways that are morally problematic (if I have argued rightly, below). If these examples go through, perhaps they point to a deeper problem. On the one hand, if the materialist turns outward, she finds only external causes cutting across human moral experience at strange and (sometimes) horrifying angles, and nothing that could either be or eventually underwrite the prescriptive authority of moral reasons. On the other, if the materialist turns inward, she finds only internal reasons unable to cut across human vagaries with any prescriptive authority and sitting in awkward juxtaposition to a materialist world of efficient causality—unexplained in their difference as reasons, even for the isolated individual. Perhaps the problem facing the materialist is not a failure of method, but of metaphysical commitments that dictate that whether the materialist turns outward or inward, what she finds is such that it cannot support the authoritative prescriptivity of moral reasons.\textsuperscript{22}

Clearly, this is an ambitious argument. Perhaps it is just wrong (though I don’t think so), and it would clearly need quite a bit of additional argumentation that will not be

\textsuperscript{22} See note 21 above for a caveat about this “inward turn”/“outward turn” language.
provided here. Perhaps the starting point would be to turn to other materialist moral philosophers and see if a pattern emerges. If such a pattern did emerge, with materialists pursuing (broadly) either an inward or an outward turn (with various hybrids) and succumbing to moral reasons arbitrariness in ways that fall out into a noticeable pattern, then it would be worth attempting to limn those patterns and see if there is an argument for why we should expect materialism in moral philosophy to lead into the coal pit of moral reasons arbitrariness.

All that, however, will have to wait for another day, as the more modest argument that three particular materialist approaches to moral philosophy lead to moral reasons arbitrariness will be plenty for one dissertation. It is to that argument that we now turn.

---

23 I have already done some related work. For my part, I think arguments similar to those advanced in the pages that follow could be successfully pressed against Jürgen Habermas and John Rawls (inward turn) and Richard Brandt (outward turn). I have pointed out some of the difficulties facing Brandt in Bradley N. Seeman, “Whose Rationality? Which Cognitive Psychotherapy? Metaphysical Facts and a Second Look at Richard Brandt’s Second Puzzle for Utilitarianism,” *International Philosophical Quarterly* 44, no. 2 (June 2004): 201-222. I have argued against Habermas in Bradley N. Seeman, “Peirce, Habermas, and Moral Absolutes,” *Journal of Interdisciplinary Studies* 12 (Summer 2000): 45-68.
CHAPTER TWO
MORAL REASONS ARBITRARINESS

This chapter develops the problem of moral reasons arbitrariness that will be the point where the later chapters put pressure on the materialist moral philosophies of Allan Gibbard, Christine Korsgaard, and John Post (and Ruth Garrett Millikan).\(^1\) In short, later chapters will argue that none of these philosophers have the resources to solve the problem of moral reasons arbitrariness. The present chapter seeks to make clear what the problem is.

As with many questions in philosophy, the question of what a reason is has proven to be quite contentious. The first section of this chapter will begin to unpack what reasons are by considering the patterns of thought that are embedded in ways people commonly speak about reasons, observing that notions of prescriptivity mark talk about reasons and that when people speak of moral reasons we also find a special notion of authority woven into the things people say. Furthermore, a consideration of the issue between philosophers who favor a causal understanding of reasons and those who favor a hermeneutic understanding reasons shows that eliminating prescriptivity and independence from reasons generally and authority from moral reasons specifically would do great violence to what people understand reasons to be. I will argue that what I call external moral reasons should be understood in terms of independent, authoritative prescriptivity where a person is able to

\(^{1}\) Note 17 in the Introduction gives my reasons for using the term “materialist” instead of “naturalist.”
recognize what is called for in a situation, though I am quick to point out that the point of the dissertation is not to argue that these authoritatively prescriptive moral reasons exist, but that moral philosophies that cannot vouchsafe such reasons run a grave risk of collapsing into moral reasons arbitrariness.

The problem of moral reasons arbitrariness is not unrelated to the ongoing debate in moral philosophy about internalism and externalism with respect to moral motivation and moral reasons. The first section concludes with an attempt to limn the essential issues in the debate about internalism and externalism as they bear on moral philosophy. Drawing on William Frankena’s classic essay on “Obligation and Motivation in Recent Moral Philosophy,” I will try to situate my argument on a map of the continuing debates. This will set some terms and orient us to the critique of the moral philosophies of Gibbard, Korsgaard, and Post.

After considering what a reason might be in the first section, the second section will approach the problem of moral reasons arbitrariness by considering a distinction between motivating moral reasons and justifying moral reasons that Christine Korsgaard (among others) has drawn. Justifying reasons are tied to the authoritative prescriptivity of external moral reasons, and some philosophers have challenged the existence of justifying reasons as somehow independent of motivating reasons. Among these philosophers is Gilbert Harman, and an examination of Harman’s objections will help clarify the problem of moral reasons arbitrariness. *Pace* Harman, it will be argued that the distinction between motivating reasons and justifying reasons is well motivated and is, indeed, crucial for our understanding of ourselves as genuinely moral beings. However, it is not crucial *in the present context* that I
show Harman is mistaken (though I think it is vitally important for moral philosophy more broadly that he is). This is because Harman’s materialism drives him to collapse justifying reasons into motivating reasons, affirming the reality of the latter, while denying that the former exist. In doing so Harman ends up being a prime example of the moral reasons arbitrariness that the present chapter will attempt to present, and I more or less endorse some of his arguments for the limited purposes of this chapter as rightly limning the way materialism shapes moral philosophy. The idea of a “justificatory reason” is meant to capture an authority that stands over against the kind of causal forces operative in motivating reasons. Once this sort of independent, authoritative prescriptivity goes missing, the choice between various motivating reasons tied to the causal forces at play in the different subjective motivational sets (SMS’s) people bring to their moral choices has no grounds for determination apart from the way those causal forces play themselves out. This is moral reasons arbitrariness, and Harman more or less embraces it. In doing so, Harman’s philosophy illustrates the problem of moral reasons arbitrariness and its relation to holding that causal forces (here exerting their force through motivating reasons tied to SMS’s) are the root of moral reasons. In short, both Harman and I agree that materialism creates problems for justificatory moral reasons, but where this leads Harman to abandon a central part of human moral experience and accept (in my terminology) moral reasons arbitrariness, it leads me to question materialism (though this dissertation does not argue that I am right in this).

The third section of Chapter 2 will strive to make the problem of moral reasons arbitrariness stand out as clearly as possible, focusing particularly on the independent, authoritative prescriptivity characteristic of genuine moral reasons. Several examples will be
given, and it will be argued that when someone believes that her moral reasons are explained by non-moral forces, this insinuates those non-moral forces into the deliberative realm where moral judgments are formed. In sum, a moral philosophy that fails to support independent, authoritative prescriptivity stands in peril of portraying morality as a field where causal forces play themselves out one way or another. The section concludes with a brief defense of the problem of moral reasons arbitrariness from an argument that could be developed from Thomas Nagel's work in moral philosophy.

So, to repeat, this chapter attempts to give reasons to believe that there is a problem of moral reasons arbitrariness, while situating the arguments to follow on a map of current debates about internalism and externalism in moral philosophy. This will open out into the central concern of the dissertation as a whole: Can some sort of rationality, as understood differently by Gibbard, Korsgaard, and Post, function within moral philosophy so as to overcome the serious problem of moral reasons arbitrariness? The main argument of the later chapters will be that the moral philosophies of Gibbard, Korsgaard, and Post all eventually succumb to the problem of moral reasons arbitrariness and the attending difficulties we saw in the Introduction.

I. What Is a Moral Reason?

A. Three Cases of Reasons-Talk

A brief look at three cases where we talk of reasons will give us a point of entry with respect to what a reason is. These cases aren't meant to be definitive or exhaustive, but they
do provide a starting point with some distinctions that we commonly mark in the way we talk about reasons. These three cases are meant to follow out a line of thinking suggested by T.M. Scanlon: “The rudimentary observation that a reason is a consideration that ‘counts in favor of’ something points toward a question, ‘In favor of what?’ and hence toward an important distinction, between those things for which reasons, in the sense I have in mind, can sensibly be offered or requested and those for which they cannot. It makes no sense to demand a reason, in this sense, for an event in the world that is unconnected with any intentional subject.” In pursuing these three cases of “reasons-talk” some important patterns emerge. To paraphrase J.L. Austin, while common usage doesn’t give us the last word, it gives us the first word. The way we talk of reasons marks some crucial distinctions that we need to attend to.

Let’s begin by considering a case where people will talk about reasons even though it is clear that reasons are not in view in the situation they are talking about. Imagine that Jim goes to the doctor and the doctor pulls out a little rubber mallet, taps his knee, and Jim’s leg jerks upward slightly. Now it is the case that we sometimes say something like this: “The reason Jim’s leg jerked upward was that the doctor tapped his knee with a rubber mallet.” Usually we speak this way if someone were to ask for an explanation of why Jim’s leg did what it did and we are speaking somewhat formally and clinically of the facts about the matter. We recount some facts of the matter in response to a demand for an explanation. More typically we might say, “Jim’s leg jerked upward because the doctor tapped it with a rubber mallet.” When speaking of the doctor tapping Jim’s knee, we would not say the

---

following, however: “The doctor gave Jim a reason to jerk his leg upward.” Instead we would say, “The doctor caused Jim’s leg to jerk upward.” Here Jim himself is bypassed, and we speak only of Jim’s leg being caused to jerk upward. No one would talk about Jim deliberating about what to do with his leg in response to the doctor tapping it, or of his weighing up considerations in favor either of jerking his leg upward or not jerking it upward.3 Jim himself, at least in his capacity to deliberate, is passive in the events that bring about his leg jerking upward. It’s something that happened to Jim; it does not involve his activity (except in his going to the doctor, not wresting the mallet from the doctor’s hand, etc.). Nor would we say that the doctor’s having tapped Jim’s knee supported Jim’s raising his leg, as we might say that Jim’s Mom’s having come down with pneumonia supported his cancellation of a business trip. There is no question of a support relationship between the tapping and the leg jerking upward.4

3 “What is the range of things for which reasons in the standard normative sense can be asked for or offered? States or occurrences that are independent of any conscious agent are excluded” (Scanlon, What We Owe, 20).

4 Paul Moser has identified at least three “support relations” relevant to reasons: causal/motivational support, evidential support, and normative support. These are tied to a “generic notion of a reason” that is specified as follows:

An item (for example, event, state, or claim) X is a reason for another item Y if and only if X is a ground, basis, means, or source of support for Y.

jerking upward. In the absence of any deliberation on Jim’s part, we do not speak of considerations that support the jerking upward of his leg.

Of course it is not the mere absence of deliberation on Jim’s part that prevents us from speaking of considerations that might support Jim acting in one way rather than another. For sometimes Jim might not deliberate about what to do, and we might still say that there were things he should have considered, that there were considerations he should have taken into account that supported a different (by implication, a right) course of action—but that he failed to do so. Thus we say: “Jim should have thought of that before acting so rashly” or “Jim should have considered what to do before he did that.” But in the case of the doctor tapping his knee, we are not tempted to say anything of the sort. Deliberation about considerations that might support one course of action over another is not in view. No one says that Jim ought to have considered the matter in a different light before jerking his knee upward. It’s not the sort of thing one can consider.

Return now to the fact that we can say, “The reason Jim’s leg jerked upward is that the doctor tapped it with a mallet.” We say this because we have been asked to give an explanation for the movement of Jim’s leg. As Fred Dretske declares: “Events are causes, but facts explain.”\(^5\) Though we might want to question this as a blanket statement, it does point to part of what is going on when we say that the reason for Jim’s leg jerking upward is that the doctor tapped it. The talk about a reason here does not pass through (so to speak) to the events in question, but remains at the level of our explanation of the events, our rehearsal of facts that we understand as coming together in such a way as to account for the

movement of the leg. We speak of reasons because of our attempt to understand the events, not because there were reasons in view in the events themselves. We understand that the facts fitting together just so is an account of what transpired with the events that occurred, and the reasons we speak of in this case are relevant to our act of seeing how this disposition of facts fits together to explain what we are interested in explaining. We can support our explanation by referring to such and such facts as considerations that must be taken into account, and by the considerations we adduce in support of our explanation we intend that others will come to see why the events unfolded the way they did. But we do not say that the considerations we appeal to in our explanation are considerations Jim took to support the jerking upward of his leg.

It is fairly obvious that our talk about reasons is ambiguous, as is our talk about something being “because of” something else. So far it is plain that when we speak of reasons, we can be speaking about a causal relation, as in the case of our effort to explain why Jim’s leg jerked upward. Neither deliberation nor considerations supporting a particular course of action are in view. Likewise, “it is clear that the word ‘because’ cannot be counted on to indicate causal connection—it can be used, for example, to indicate the relation between premises and conclusion of a valid argument, which is surely not a causal relation.”

Thus, following Kieran Setiya, who disambiguates the word “because” with the notation because\(_a\) to indicate the sense of because where it indicates a reason, we could

---

denote the use of “reason” to talk about an explanation in terms of causes with reason$_{CE}$. This notation will not be employed here, but we need to make plain that in our ways of talking about events like Jim’s leg jerking upward when the doctor taps his knee there is a clear sense that Jim does not deliberate about considerations in favor of the upward jerk of his leg, and the doctor’s tapping his knee does not support what his leg does, but causes it. Our occasional reasons-talk in cases like this does not pass through to the events themselves, but is operative within our interest of understanding the event with an explanation worked out in terms of considerations about the relevant facts.

A very different case of reasons-talk than the one we have been considering is the case of Jane working through a geometry problem. At a particular point in the course of working the problem, Jane employs the Pythagorean Theorem to solve the problem successfully. We might say something along the following lines: “Jane saw that she had come to a step in the problem where she had a reason to use the Pythagorean Theorem.” Though perfectly in order, that sounds a bit stilted; instead we might often say, “Jane saw that she needed to use the Pythagorean Theorem.” And a teacher grading Jane’s (unsuccessful) attempt to solve the problem says, “Jane needed to use the Pythagorean Theorem here, but she failed to see it.” Upon reading her teacher’s comments, Jane might say to herself, “Oh, I get it now. Why didn’t I see that one ought to use the Pythagorean Theorem there? It’s obvious, now that I see it.” Or if the solution is dawning slowly, she might say: “Oh, it’s coming to me at last. I should have used the Pythagorean Theorem there.”

---

Here the reasons-talk marks some sort of difference from the case of Jim’s knee jerking upward. For one, Jane is clearly not being bypassed. Where we speak of Jim’s knee in the one case, here we speak of Jane herself. Jane is active, and it is she that has the reason and responds to it by using (or failing to use) the Pythagorean Theorem. Jane, and her activity in deliberating about the problem and coming to see what should be done, is essential in a way Jim was not when the doctor tapped his knee. Jim is a spectator passively observing something happening to him, while Jane is an active and essential agent in the working of the geometry problem. Indeed, Jim could be unconscious when the doctor taps his knee and his leg would still jerk upward; there is no question of that with Jane and the geometry problem. Her activity of deliberating, weighing, considering is essential. Yet there is also a sense of recognizing something that is there apart from her activity; as she works the problem the solution dawns on her, and she might say that “It was staring me in the face all along. Of course, I need to use the Pythagorean Theorem.” There is a strong sense of “coming to see” what is there, whether she recognizes it or not. This is the place where the Pythagorean Theorem is appropriate. Jane has a sense of recognizing a support relation that is there independently of herself, and we can say that she takes the reason to be external to herself in some sense. “It is the right thing to do here,” Jane tells herself as she sees that it is fitting or called for. And if no one had ever worked that problem before, the sense of coming to see things rightly would be undiminished.

As Jane works through the geometry problem and sees that she ought to use the Pythagorean Theorem she experiences a kind of normativity, a requirement that she experiences as uniquely right or fitting. As she deliberates about the problem, she sees
considerations that lead toward a specific conclusion or a next step that needs to be taken, and that the failure to take this step would be just that: a failure. Jane is aware of the possibility of getting things wrong or getting them right—there is a possibility of error facing her, a possibility that will become reality if she doesn’t do the right thing with respect to the problem. As she works through the problem, she seeks to “get it right” and “find the solution” that is supported by the reasons she is seeking to ferret out. She seeks to do what needs to be done or should be done or would be right to do, and she experiences this as and talks about it as a matter of recognizing, coming to see, or identifying what ought to be done. She is seeking to do what she has reasons to do. If all goes well, she works it the way she does because of the reasons she sees. And when she sees that she has reason to use the Pythagorean Theorem, Jane “is aware of the reason as a reason; she identifies the good-making properties of the action under the description ‘good’ or ‘reason’ or ‘right,’ or some such normative description. She does not act merely in accordance with a normative consideration but on one.” Jane recognizes and rightly acts on a support relation that she comes to see as standing over against her and indicating what should be done: the Pythagorean Theorem is uniquely called for here, Jane sees. This is what Jane recognizes and rightly acts on. In this, Jane’s experience is wholly unlike Jim’s experience when the doctor taps his knee with a rubber mallet, where Jim neither recognizes nor acts on anything. Something simply happens to Jim. Nothing whatever is called for.

A third and final case strikes us as different yet again. Suppose it’s a windy day and Ted is leaving the library. When Ted opens his car door, the wind whips it out of his hand.

---

and smashes it into the shiny, new Acura parked next to his car, leaving a huge gash in the door. As he looks at the gash, Ted “goes back and forth” with himself about what to do:

“Should I leave a note with my contact information on the windshield or just take off?”

Considering Ted’s situation, we might say, “Ted has a reason to get out of there, but he has a better reason to leave a note.” Or, more naturally, “Ted should leave a note, even if he doesn’t want to.” And Ted himself might think along the following lines as he considers what to do: “On the one hand, if I take off without leaving a note, I will save some money that I will have to pay and no one will be the wiser. But, on the other hand, leaving a note is the right thing to do.” If he leaves the note, he will likely think, “Well, at least I did the right thing,” even if he is more than a little annoyed about the whole situation. If, however, he takes off without leaving a note, he may be happy about saving the money and may soon put the whole thing out of his mind, but it is likely that he will feel at least a little guilty in the short term and he might say to himself: “I did something wrong,” or “I disobeyed my conscience.” Certainly the owner of the car (in a case of what theologians call “judicial sentiment”) will say, “Whoever did this was wrong to leave without giving me a way to contact her.”

This third case is closer to that of Jane and the geometry problem than it is to Jim’s knee being tapped by the doctor. Unlike the latter case, our typical ways of talking about Jane and Ted do not bypass the person. Both Ted and Jane are considering what to do and deciding what course of action is supported by the reasons. Moreover, both Ted and Jane have a clear sense that a matter of being right or wrong is at stake; there is a normative issue.

---

9 I have actually had this situation happen.
There is a clear issue of “ought” that Ted and Jane are alive to, and they are trying to be responsive to the reasons that they believe to inhabit the situation. They are alive to the fact that something is called for in their situation, a judgment or action that is uniquely supported and may be recognized as such. Jim has no sense of this at all as the doctor taps his knee. For Jim, unlike Jane and Ted, there is no activity of trying to weigh considerations that support one course of action over another, or of anything that ought to be done or of an error that could be made with respect to what should be done in the situation. And Jim certainly has no sense of identifying something right, an experience common to both Jane and Ted.

Yet there are also significant differences between Jane laboring on her geometry problem and Ted debating about what he should do about the gash in the car door. Chief among these is that in Jane’s experience there is no sense of conflict, while this is a prominent feature of Ted’s experience. When Jane sees that the situation calls for the Pythagorean Theorem—that this is what she should do—she immediately sets about doing so. There is no hesitation, no resistance to be overcome, no conflict. A case of geometrical akrasia would be strange indeed. There is no experience of, “On the one hand, I see that I should use the Pythagorean Theorem; but, on the other hand…” No. Once she sees that this is what ought to be done, she just gets on with it.

But in Ted’s case things are different. The difference is captured in what we are most naturally inclined to say: “Ted should leave a note, even if he doesn’t want to.” It seems to be a common human experience in moral situations to sense opposition between what “should” be done and what one “wants” to do, with a sense of authority attending the
“should” and standing over against an attraction or a drive attending the “want.” There is the sense of something authoritative, commanding—even imperious—in the sense of should that Ted experiences that is not present with the geometry problem. In both cases we can “fail”; only in the moral case can we “violate” or “rebel” in the sense of resisting what we think should be done. Jane is not tempted to resist or rebel against the realization that she should use the Pythagorean Theorem, while Ted may be sorely tempted indeed, perhaps to the point of defying the claim that he takes himself to be under and refusing to leave a note. The two kinds of reasons operative in these different situations, epistemic reasons for belief (and attending dispositions to work the problem in the specified way) and moral reasons for action, differ greatly in this respect. When Jean Hampton notes that “human beings can take (highly) negative attitudes toward what they know to be their reasons,” she is clearly talking about moral reasons. In moral situations, “reasons feel like orders”—a fact indicated by the sense of rebellion that people sometimes have with respect to them.

10 It is true that we can speak of “violating” the principles of geometry, but there is no sense of willful rebellion or of following some other, more attractive consideration. There is more a sense of having failed to see what should be done rather than of seeing and choosing not to. We would have to tell some special story to make sense of violating (in the sense of rebelling against) the laws of geometry. Jane’s tormenters are trying to brainwash her, and in an act of defiance she willfully violates the laws of geometry to show her independence. Such a case clearly need not concern us here.

11 The distinction here has been debated, but is fairly common in the literature. It does seem to me that some sort of rough and ready distinction can be discerned in attending to examples like the ones being canvassed here.

12 Jean E. Hampton, The Authority of Reason (Cambridge: Cambridge University Press, 1998), 89-90. I take it that Hampton’s interest in making clear this distinction underlies her clear separation of prescriptivity and justification (87-90).

13 Hampton, Authority, 106.
The difference resides in two kinds of considerations that show up in the statement that “Ted should leave a note, even if he doesn’t want to” and in the familiar “On the one hand… on the other hand…” structure of debates that often play out in deliberations about what to do in moral situations. On the one hand, there is a want or desire that the person finds attractive, and this want or desire is often experienced as a temptation: It would be nice not to leave a note on the windshield, since I would rather use the money I will need to pay in some other way and it will take time to interact with the owner of the car, and those interactions will take time I could use otherwise and may be unpleasant, and so forth. These wants can be experienced as attractions or drives or forces welling up from within, and they pit themselves against something that feels very different. And so, on the other hand, the person in some way runs into what should be done, and, not to put too fine a point on it, there is a sense of needing to obey. It is hard to know how to put this sense, but it feels like what should be done overrides or trumps other considerations, or that it commands or rightfully claims obedience. And there can be a sense of deserved guilt if one fails to live up to the rightful claim that comes to expression in a person’s moral reasons. This again is in strong contrast to the geometry problem, where Jane might feel inept if she failed to use the Pythagorean Theorem where she should have, but she would not feel guilty (except, perhaps, if she neglected to study). In Ted’s situation, on the other hand, there is a sense of a rightful claim upon his actions. If he chooses to take off without leaving a note, he may well feel guilty for doing what he knows or believes to be wrong. And even if he fails to recognize the claim, others may well say that he should have done otherwise and people

---

14 Of course on some happy occasions they coincide with what the person sees should be done, but there are other times when there is internal conflict.
would believe that he remains accountable to reasons that he should have recognized and heeded, even where he failed to do so.

At this point we are clearly a long way from Jim’s leg jerking upward as the doctor taps on his knee. I take the case of Jim’s knee to be a paradigm and unproblematic instance of causation, however troubled our philosophical attempts of formulating a clear conception of causality may be in the wake of Hume’s challenges.\textsuperscript{15} Quite different are Jane and her geometry problem, and Ted debating about what he will do about the gash he made (even if only in not holding onto the door securely on a windy day) in someone else’s shiny, new car. In these cases we are dealing with reasons.\textsuperscript{16} Jane and Ted decide what to do, and people do not bypass their deliberative activity in talking about the situations they are in. Jane and Ted are alive to certain support relations present in the situation and are trying to discern them. In Korsgaard’s phrase, both Jane and Ted are acting on normative considerations after having worked to identify them.\textsuperscript{17} But in Ted’s case alone there is a sense of something that rightfully claims his obedience and there are temptations also to evade or defy the reasons for doing what he should do. Both Ted and Jane are in normative situations, but Ted’s situation involves moral normativity that is distinguished by a possibility of resistance alien

\textsuperscript{15} As Hampton puts it, “Does ‘structure’ refer to a certain kind of causal explanation in science that is not possible in objectivist moral theories? Perhaps, but it is an idea that is extremely hard to develop satisfactorily. On the face of it, an appeal to causes ought to be troubling to the naturalist by virtue of Hume’s attack on the idea” (Hampton, Authority, 41). Hampton herself refers to Wesley Salmon, “Four Decades of Scientific Explanation,” in Scientific Explanation, ed. Philip Kitcher and Wesley Salmon (Minneapolis: University of Minnesota Press, 1989).

\textsuperscript{16} T.M. Scanlon makes a helpful observation in this connection:

\textsuperscript{17} There is some recognitional element involved in acting for a reason, some sense that any causality involved in reasons passing through intelligibility. See Robert Audi, Action, Intention, and Reason (Ithaca, NY: Cornell University Press, 1993), 15-17. Audi’s own account of reasons is causal. I have some hesitations about this, though Audi’s own causal account of reasons incorporates elements that address some of my hesitations. I will canvass some of these issues below.
to working a geometry problem—a resistance directed toward what is perceived as some sort of authoritative claim upon his action. This authoritative claim has been variously described (across a wide swath of philosophical views) as “an ineliminable prescriptive, or imperative, component,” as having “normative authority,” as a kind of authoritative “choice-worthiness,” as having a kind of “authority” or “normative force,” as being an instance of “second-personal authority,” as being “intrinsically prescriptive” or “objectively prescriptive,” and as “directive authority,” “objective authority,” or “culture-independent prescriptivity.” Each of these ways of trying to capture this central part of human moral experience moves in a helpful direction.

Taking my cue from these ideas, I will refer to this central feature of moral reasons as authoritative prescriptivity, though I do this with some trepidation. By prescriptivity I mean


19 Carson, *Value and the Good Life*, 206; see also 251.


21 Scanlon, *What We Owe*, 55, 57.


25 Mackie (*Ethics*, 38-9) also uses this phrase. I largely find Mackie’s argument in the first chapter of *Ethics* convincing, though where he runs it in the direction of *modus ponens* I would run it in the direction of *modus tollens*. My hesitation in talking about “authoritative prescriptivity” comes from the non-cognitive heritage of “prescriptivism” running through R.M. Hare (for example, R.M. Hare, *The Language of Morals* (Oxford: Clarendon Press, 1952; reprint, 1961)), though, as will emerge later, I do believe there is a very important element of endorsement in ethics. That said, Thomas Carson puts the same worry I have quite well: “On [R.M.] Hare’s view, saying that x is better than y does not imply that it is correct to prefer x to y. Because of this, Hare’s theory cannot provide an adequate account of the nature of normative disagreement or the causal efficacy of normative judgments. Suppose that two people disagree about whether x is better than y. When we disagree about whether x is better than y, we are not just prescribing different choices for this and
to pick out an aspect that is a familiar part of common experience, that experience like Ted’s being faced with what he ought to do—the experience of being directed, where I take it that this has a recognitional element; Ted and Jane see what ought to be done, they are aware of the directive force in their respective situations, and if they do it they are acting on the reasons they have. They are responding to features in their situations that direct them, and they act on considerations that support their action because they are aware of that support relation and want to do what there is reason to do. They recognize a support relation that stands over against them, independent of that moment of recognition. As Jean Hampton observes, “we are moved to act from moral reasons because we understand their objective authority.” But this applies not only to moral reasons, but also to the sort of reasons Jane is alive to in working the geometry problem. In the reasons Jane and Ted have, a directive element comes to expression and is seen as directive, and it is through that recognition of the directive force of the reasons that the reasons move people. In short, they see that a particular judgment or action is uniquely called for and that the support for this is independent of their own desires.

relevantly similar cases; we are claiming that our own preferences are correct and the other person’s mistaken,” (Carson, Value and the Good Life, 178). That strikes me as exactly right. R.M. Hare’s son, John Hare, provides a salutary corrective here in talking about what he calls “prescriptive realism”: when we commend something “we are not putting something into the world that was not there before… we are responding to something that is already there,” (John E. Hare, God’s Call: Moral Realism, God’s Commands, and Human Autonomy (Grand Rapids, MI: Eerdmans, 2001), 39). I think Hare is misguided in trying to work out this prescriptive realism in terms of Kantian morality, and I have definite disagreements with him as he makes this effort; but his prescriptive realism, stripped of the misbegotten attempt to appropriate Kant, is a helpful model.

26 Hampton, Authority of Reason, 105; emphasis in original.

27 There is, of course, a fuller picture here, one where we would also need to wade into the thickets surrounding how and whether wants are essential to motivation. The majority position is decidedly Humean, but there are interesting questions pressed by John McDowell and others. I don’t need to hack my way into these dense brambles here.
But there is a dimension to Ted’s moral situation that is absent in Jane’s geometrical ruminations, and it is this dimension that I mean to point out by talking about *authoritative* prescriptivity. In short, there is a sense of being commanded or being expected to submit that is missing in geometry. In both cases there is a clear sense that there is something that is right to do; but only in the moral realm is there a sense of a *binding* imperative. Only in the moral case is there something that *must* be done, and a sense that one is in fact *guilty* if one does otherwise; one has disobeyed. So, while there is a clear sense of failure to which both Ted and Jane are alive in their respective situations, only in Ted’s case does that failure have a sense of rebellion. That sense is missing in the geometry case; once both Ted and Jane see what is called for in their respective situations, only Ted is tempted to rebel against what he sees should be done. That against which Ted is tempted to rebel is the authoritative element unique to moral reasons. As Paul Moser puts the point, we are all faced with “Love’s Demand” and “we consistently fail” to live up to this demand. “Our conscience, if unsuppressed, convicts us of this failure, and we then experience guilt and even shame.”

There is a familiar commandingness in moral questions that we can “violate” or “flout”—and it is that to which I am referring to when I speak of *authoritative* prescriptivity. Whatever we call it, this independent, authoritative prescriptivity is a clear feature of genuine moral reasons, even if—like the notion of causality—it is notoriously difficult capture the experience in an adequate theoretical construct.

---


29 I believe Jean Hampton overstates the difficulties when she says that “the idea of an authority that is objective is ineffable—that is, impossible to pin down in a way that seems to make sense,” *Authority of*
kind of independent, authoritative prescriptivity that emerges in the three examples of how people talk and think about reasons as opposed to causes will be understood to be the core of what will be called external moral reasons, with (1) “external” referring to the independence of these reasons from the subjective motivational sets of contingent personal beings, (2) “reason” referring to the called-for-ness or prescriptivity that is recognized in the support relation (as seen in the cases of Jane and Ted recognizing and acting on an instance of such support as uniquely calling for a particular judgment or action, as opposed to the very different case of Jim’s leg being caused to jerk upward), and (3) “moral” as referring to the sense of authority or justifiable command or even rightful demand for obedience that is unique to moral reasons (as seen in cases relevantly like Ted’s, and unlike Jane’s and Jim’s).

Thus we may understand external moral reasons, as will be referred to in what follows.

It will not be the business of this dissertation to argue that such reasons do in fact exist (though I believe they do), but to suggest that moral philosophy may be in grave trouble if they do not by looking at three contemporary materialist moral philosophies that fail to support external moral reasons and arguing that they each succumb to moral reasons arbitrariness. But before these moral philosophies can be profitably examined, several

---

Reason, 99). Indeed, it seems to me that Hampton turns away from the most hopeful avenue for explicating “authority” when she argues that personhood cannot be made central to the objective authority she seeks to understand (Authority of Reason, 106-108). I do not find Hampton’s arguments on this point at all convincing, but this issue will not be taken up here. Having thus hamstrung her own efforts, Hampton is forced to conclude that the authority she is rightly concerned with is “ineffable.”

In a not unrelated vein, T.M. Scanlon simply starts his moral philosophy with the notion of a reason as basic: “I will take the idea of a reason as primitive. Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favor of it. ‘Counts in favor how?’ one might ask. ‘By providing a reason for it’ seems to be the only answer,” (Scanlon, What We Owe, 17). Scanlon does not have the notion of authority strictly in view here, but more the prescriptivity of reasons that operates through a person’s recognition of support relations. While the notion of a reason is thus not wholly clear, Scanlon argues that “desire is not a clearer notion in terms of which the idea of having a reason might be understood,” (7). Again, though, in addition to this “sharing the guilt” strategy, perhaps reasons could be worked out in terms of personhood—though that will not be undertaken here. Stephen Darwall, The Second-Person Standpoint, starts down some interesting paths in this connection.
essential tasks remain. The first of these is to examine the commonplace but difficult distinction between reasons and causes by looking at the way some materialists have dealt with reasons and then considering Robert Audi’s contrasting approach. The distinction between reasons and causes features prominently in the discussion in succeeding chapters (especially Chapter 5), so we now turn to the task of how that distinction may be drawn. The question is what role, if any, a hermeneutic or recognitional dimension plays in distinguishing reasons from causes.

B. Reasons and Causes

1. Materialists on reasons. The relationship between reasons and causes is a tortured one, in no small measure because philosophers have not succeeded in making either notion perspicuous, and yet we cannot do away with either notion, since they are both central to the understanding of everyday experience. Clearly we think about many events in terms of causal relations of a non-personal sort, such as one billiard ball vectoring toward the corner pocket when struck just so, or the upward jerk of one’s leg when one’s knee is tapped with a rubber mallet. At the same time, there are instances where people understand themselves to believe or act on the basis of their recognition of support relations that uniquely uphold one particular belief or action as what ought to be believed or done. There is, of course, much interest today in making the latter beholden to the former so that the non-personal may be seen to be more fundamental than the personal, thus avoiding anything “spooky” or “mysterious.”

30 The aim of avoiding “mystery” is a major theme in one of the modern classics of action theory, Donald Davidson’s, “Actions, Reasons, and Causes”: “I would urge that, failing a satisfactory alternative, the
few] are interested in the extent to which human agents are part of ‘nature’. And there is “much interest in whether any single conception of explanation can account for all explanations of empirical phenomena, or at least for both scientific explanations and the most important kinds of commonsense explanations of action.” Yet there remains the stubborn fact that human beings act on the basis of recognizing what should or ought to be done, and even those who would firmly subordinate reasons to a single conception of explanation grounded in efficient causality find themselves compelled to account for the fact that people describe their actions as warranted and understand their acts to have arisen from their recognition of that warrant. As Joseph Raz has put the matter, agents undertake their actions and explain them in light of “a story which shows what about the situation or action make it, the action, an intelligible object of choice for the agent, given who he is and how he saw things at the time.”

A rather different view is presented by Fred Dretske in his essay “Reasons and Causes.” Dretske opens by declaring, “I am a materialist who thinks that we sometimes do things because of what we believe and want. I pretty much have to accept the idea, then,

---


32 Audi, “Wants and Intentions,” 123.

33 Joseph Raz, “Agency, Reason, and the Good,” in Engaging Reason (Oxford: Oxford University Press, 1999), 24; quoted in Setiya, Reasons without Rationalism, 62. Setiya, it should be noted, is advancing a causal notion of reasons that is set against Raz’s understanding of reasons and, indeed, any normative notion of reasons that is operating “under the guise of the good.”
that reasons are causes.” The project Dretske sets for himself in this essay is a materialist explanation of how it could be that “the semantic aspect of reasons, the what-it-is we believe and desire, is the property that, in addition to rationalizing or justifying what we do, also figures essentially in the causal explanation of what we do.” As Dretske sees, in any understanding of reasons that would not do violence to the very idea of acting for reasons, the semantic properties central to the agent herself must be seen to do real explanatory work somehow—and this is not easily brought within the ambit of materialism. After canvassing a few materialist accounts of reasons that strip semantic properties of any explanatory work, Dretske writes, “I find this result quite unacceptable. It implies that what we believe, intend, and desire has no bearing on what we do. It implies that what a person thinks has as much relevance to what he does as what a sound means has to the amount of pressure it exerts on a glass…. If the semantic properties of reasons, the what-it-is we believe and desire, is irrelevant to explaining their causal properties, what it is they make us do, then the fact that they are causes, taken by itself, is or should be very little solace indeed.”

Looking at Stephen Stich’s biological account of reasons, Dretske complains that “the upshot of such arguments is that we may believe things and we may desire things. And our beliefs and desires may even cause us to behave in certain ways. But nothing about what we believe and desire, not even (it seems) the fact that we believe and desire something, is pertinent to

---

35 Dretske, “Reasons and Causes,” 1; emphasis in original.
36 Dretske, “Reasons and Causes,” 3; emphasis in original.
understanding *why* we have this way… I find this conclusion unacceptable.”

Dretske is right here, and, having seen this, he sets about a materialist vindication of the recognitional aspect of people acting on their understanding of a support relation—an understanding wholly missing from Jim’s leg jerking upward, but essential to Jane seeing that the geometry problem calls for the Pythagorean Theorem and Ted understanding that he should leave a note on the windshield.

Dretske’s attempt to provide this account falls well short. Dretske’s first move is to put what he calls “The Design Problem” at the center of this issue, where this has to do with “simple learning situations: How does one get the pupil to say ‘oak’ (not ‘maple’ or ‘pine’) when shown the distinctive markings of an oak tree?”

“The Design Problem” is thus essentially a matter of how to make an indicator trip a switch. Dretske argues that “for any system $S$ for which The Design Problem has been solved, we have some internal state or condition in $S$ that indicates, means (in Grice’s natural sense of meaning), or represents (in, I think, one sense of this word) something about how things stand outside of $S$.” This is part of Dretske’s “tactical detour” of trying to approach the targeted semantic properties through a certain kind of relational properties. But, as Dretske admits, even if his tack succeeds, he “will not yet have shown that behavior… can be explained by the fact that a

---

37 Dretske, “Reasons and Causes,” 5; emphasis in original.

38 Dretske, “Reasons and Causes,” 9. Whether solving this problem would make any headway on the semantic problem is very much open to question, so even this formulation of the problem is questionable. But let that pass.

39 Dretske, “Reasons and Causes,” 9; emphasis in original.

state has certain meaning-constitutive properties.”41 In this he hopes that he will have done enough “to make my near miss intriguing.”42 But, the sense of “meaning” in his “near miss” is indeed “a pretty anemic sort of meaning, not rich or intentionally robust enough to serve as the propositional content of a belief or a desire.”43 For, of course, a thermostat can have the sort of meaning that is in view in the relationship Dretske has laid out, and there is no question of a thermostat doing anything like acting on the recognition of the semantic properties of the situation. A bimetallic strip bends as the kinetic activity of the molecules changes with a change in temperature and the strip trips a switch that starts the furnace. Clearly we are in the realm of Jim’s knee jerking upward when the doctor taps it with a rubber mallet, and still a good way from Jane seeing that she needs to apply the Pythagorean Theorem at this point in the geometry problem (no closer than when Dretske started, really). At the end of his “tactical detour,” Dretske observes that “though suggestive” (as he hopes), this “tendentious description of the thermostat ignores our involvement in the proceedings.”44 But this is just what cannot be done. As was noted in talking about Jim’s leg jerking upward when the doctor taps his knee, it is characteristic of cases of causation like this that Jim’s involvement in the proceedings is bypassed. Dretske’s sketch leaves out everything that is essential, and he makes little if any progress toward a materialist account of how the semantic properties carry real explanatory weight.

44 Dretske, “Reasons and Causes,” 11; emphasis in original.
However problematic his account of the semantic properties of reasons may be, Dretske at least recognizes the need for such an account and attempts to provide one. Other materialists don’t even bother with the attempt. Thus, Rüdiger Bittner declares that “the fundamental pattern of doing things for reasons is: given this, somebody does that; and in such a story the agent’s desire does not figure… Thus the story of somebody’s doing something for a reason is not a psychological one, uncovering the inner springs of the action. It is a historical one, locating the action among things happening.”45 Causal forces pass through the agent and determine her action, the semantic properties perhaps tagging along sometimes. Thus the men in the classic experiment who picked out a picture of one woman as more attractive than another picture of the same woman—the only difference between the pictures being that in one the woman’s pupils had been airbrushed to appear more dilated (a sign of sexual arousal)—were acting for reasons no less for the fact that they were completely unaware of why they chose the one picture over the other.46 “Rational agents are animals sniffing their way through the world. They are not in control. They are given to what they encounter.”47 Once again, though more baldly than in Dretske’s materialist philosophy of reasons, the activity of the agent is bypassed: “being active does not lie at the heart of agency.”48 “The picture of rational agents emerging here may be summed up by saying that they are worldly creatures through and through. Taking their clue from what they encounter and unguided by any authority or law of independent

46 Bittner does not use this example.
47 Bittner, Doing Things for Reasons, 164.
standing, they simply continue the threads of the world.” Human beings click along the series of tracks and switches set for us by the world, falling to one temptation or another, our agency consisting of nothing more than that. On Bittner’s account, acting for reasons has nothing to do with recognizing the meaning of and responding to prescriptivity.

Bittner’s attack on reasons goes deeper still, making plain that he also—perhaps particularly—targets the authority of reasons as something that cannot be sustained. “We are not by authority of reasons called upon to do anything.” When one feels the demand for obedience expressed in a reason, one can remember that this, too, is only a temptation of the world, and one may not wish to fall into it. “Rational agents, then, are characterized on the present account by the absence of inner domination and of normativity. Rational agents are in this sense, free: not subject to master or law within.” Bittner is no doubt right when he says that “it may even be an attraction of this picture of rational agents that it does away with the notion of a special dignity they carry. Dignity is something one has to live up to. Thus it is based on pressure, and one has to be in arms, at least against oneself, to protect it. There is, by contrast, relief in thinking that nothing particular is expected from us by virtue of our rational nature. There is relief in being undignified, in shedding any higher calling, in

49 Bittner, Doing Things for Reasons, 170.

50 “Falling to temptation is just what, on the present conception, rational agency amounts to,” (Bittner, Doing Things for Reasons, 164.)

51 Bittner, Doing Things for Reasons, 169.

52 Bittner, Doing Things for Reasons, 170.
being unreservedly worldly. 53 Indeed there is (I will refrain here from drawing contrasts with Socrates dissatisfied).

I have no interest in arguing against Bittner here; indeed, I’m inclined to think there is a certain consistency working itself out in his thinking (though I will make no attempt to argue the point here). The point of this brief consideration of Bittner and Dretske on reasons has not been to argue against their account, but rather to throw into bold relief the vital importance of the recognitional or semantic element of reasons. Theories of reasons that fail to give a convincing account of this element end up leaving our “reasons” at the mercy of efficient causes that potentially cut across what have seemed to many to be the moral reasons we have at some bizarre angles. (This dissertation is not an attempt to argue that point generally, but to illustrate it at work the work of three philosophers; for all I say here, perhaps some other account succeeds where they fail.) There is a normative element of reasons that passes through and operates irreducibly in a hermeneutics of the situation to which the agent is alive. If the element of a prescriptivity to which an agent responds to as such cannot be vindicated, then human beings don’t have reasons in the sense many (all?) of us have thought we have. Some account of the legitimacy of the authority many human beings take to be expressed in moral reasons is also essential if human moral life is to be more than an “error.”

Before moving on to consider the recognitional element of reasons by engaging Robert Audi’s work on reason, the vital points illustrated in our quick look at Dretske and

53 Bittner, Doing Things for Reasons, 172. Perhaps my reading of Foucault is off base, but there seems to me to be a deep affinity with a central strand of Foucault’s thought here.
Bittner struggling with reasons can be reinforced by glancing at T.M. Scanlon’s very different materialist account of reasons.

As mentioned above, Scanlon takes reasons to be “primitive.” “I do not believe,” Scanlon declares, “that we should regard the idea of a reason as mysterious, or as one that needs, or can be given, a philosophical explanation in terms of some other, more basic notion. In particular, the idea of a reason should not be thought to present metaphysical or epistemological difficulties that render it suspect.” Scanlon

This is not an entirely tension-free position for a materialist to take, but it does take the notion of a reason seriously. Thus freed of the burden of showing how reasons fit into a materialist ontology, Scanlon begins to develop an account of what he calls “reasons in the standard normative sense.” Scanlon takes these reasons to have “authority” and “normative force which resists identification with any proposition about the natural world.” Scanlon makes clear that he is concerned not with the fact that someone takes herself to have a reason, “but with whether it is a good reason—a consideration that really counts in favor of the thing in question.” And in light of the fact that there are normative standards applicable to one’s actions, intentions, and beliefs, “one can properly be asked to defend these attitudes according to the canons

---

54 Scanlon, What We Owe, 3

55 Indeed, Scanlon’s high-pitched insistence that materialist metaphysics present no problem for reasons might of itself give one pause.

56 Scanlon, What We Owe, 20.

57 Scanlon, What We Owe, 6, 55.

58 Scanlon, What We Owe, 57.

59 Scanlon, What We Owe, 19; emphasis in original.
relevant to them.”

Agents can understand the relevance of certain canons by which their actions may rightly be judged, and they are expected to be able to defend their actions by reference to those canons—canons which are taken to have authority. And, with respect to these canons, one can be mistaken or may rightly take a change in behavior to be a correction. And unlike Bittner, who insists that “doing things for reasons is not a privilege of humans (and, possibly, higher beings), but shared with all sorts of lowly animals,” Scanlon argues that human “reflective capacities set us apart from creatures who, although they can act purposefully, as my cat does when she tried to get into the cabinet where the cat food is kept, cannot raise or answer the questions whether a given purpose provides adequate reason for action. We have this capacity, and consequently every action that we take with even a minimum of deliberation about what to do reflects a judgment that a certain reason is worth acting on.”

Now, I have differences with Scanlon on some of these points, and the significant points where we agree I would flesh out quite differently; but he clearly takes reasons very seriously and makes points that take a kind of prescriptivity and even a certain authority seriously. There is much that he has right.

There is a problem, however. To take reasons as “primitive,” of course, is just to insist that reasons do fit in a materialist world after all; but it is not to provide any account of how this might work. Scanlon can declare that all this about correct and incorrect moral judgments justified by reasons “should not strike us as mysterious,” but he stands opposed

60 Scanlon, What We Owe, 22.

61 Bittner, Doing Things for Reasons, 171. Later on the same page Bittner declares that “‘Humanity’ is a term lacking all honorific overtones and moral significance; it is a term like ‘anthood’.”

62 Scanlon, What We Owe, 23.
to a hoary line of materialists who see nothing less than a “mystery”—if not something
downright queer. Some of these materialists are willing, even happy, to dismiss any notion
of authoritative prescriptivity, while others, Scanlon among them, find it worrisome “that
there may be no right answer to questions of right and wrong. This is a serious challenge,
and it may seem that in order to answer it we must provide a metaphysical account of the
subject matter of judgments of this kind. I believe that this is not what is necessary,
however. The question at issue is not a metaphysical one.” Thus Scanlon believes that if
only we can characterize “the method of reasoning through which we arrive at judgments of
right and wrong,” then “no interesting question would remain about the ontology of
morals.” Yet for all of that, the notion of a reason at the heart of Scanlon’s procedure is
taken to be primitive, and—as Dretske’s machinations make plain—it is not clear how the
materialist can make reasons fit into the materialist world, Scanlon’s assertion that we need
not worry about it notwithstanding. It does not help to assert, as Scanlon does, that what
“is special about reasons is not the ontological category of things that can be reasons, but
rather the status of being a reason, that is to say, of counting in favor of some judgment-
sensitive attitude.” That there need to be “standards for arriving at conclusions about

63 Scanlon, What We Owe, 4.

64 Scanlon, What We Owe, 2.

65 Scanlon, What We Owe, 2. Scanlon would say that moral facts are grounded in what we can justify
to each other and that metaphysical matters are adiaphorous. Whether Scanlon smuggles in materialistically
suspect normative contraband in his notion of “justify” and just how well he can solve difficulties of justifying
what to which group of people in a way that is metaphysically austere enough to count as materialism would be
a question worth pursuing, but one that would take us too far afield here. I believe Scanlon’s account falls well
short, but arguing this would be a lengthy detour. In this brief overview, raising a few pointed questions is
sufficient for the purpose at hand.

66 Scanlon, What We Owe, 56.
reasons” in the sense of governing the procedure does not soften the tough questions about how that could be made to fit in a materialist world. Scanlon has not rendered some project like Dretske’s unnecessary.

Where Bittner more or less rejects any vestige of authoritative prescriptivity in reasons and Dretske struggles to justify even a very watered-down notion of such prescriptivity, Scanlon rightly sees that some such notion is essential to ethics and wisely embraces it. Whatever traction his moral philosophy gets owes to this embrace. But what goes missing is any sense of how his system fits into a materialist ontology. He simply assures us that it does. In contrast, Bittner just embraces the ontology and dismisses anything authoritatively prescriptive in reasons, while Dretske struggles to effect some sort of rapprochement—without notable success (though I have more pointed to that failure than I have argued it at length). All three of these materialist philosophers point in their different ways to the importance of retaining an authoritative prescriptivity that comes to expression in moral reasons and is recognized by agents and responded to as such; and each of them also show in their different ways the difficulty of making any of that fit with materialism.

One way of making clear the difficulty materialists have with reasons is to emphasize the distinction between reasons and causes and to show what it is about reasons that materialists struggle to capture—as, indeed, has already begun to emerge. Turning now to the reasons/causes distinction, emphasis will be laid on the importance of understanding or recognizing what is called for or supported in a reason, so that one judges or acts on that
recognition of support. The distinction pivots on the hermeneutic dimension essential to reasons, as will emerge through careful attention to Robert Audi’s helpful work on reasons.

2. Reasons and recognition. In my emphasis on the importance of recognizing an authoritative prescriptivity that comes to expression in reasons, I have clearly been coming down on the side of what is sometimes called a “hermeneutic” understanding of reasons rather than the “causal” account favored by Donald Davidson, Kieran Setiya, and many others. The division between these two accounts of reasons is summarized by Christine Korsgaard in a passage notable for its clarity:

Empiricists tend to think that reasons are provided by our mental states, especially our desires; that the relevant facts concern the desirability of the goals to be achieved through action; and that the relation between reasons and actions is causal. Rationalists tend to think that reasons are provided by the facts in virtue of which the action is good, that these facts need not be limited to the desirability of the goals that are achieved through action, but may concern intrinsic properties of the action itself; and that the action is caused not by the reason, but rather by the agent’s response to the reason. 67

In the terms as they are laid out by Korsgaard, I do favor the “rationalist” account, though I have reservations about the way rationalists work out the points in question (and would not be happy with the label, in any case). However, there are important insights on the “empiricist” reading of reasons as well. Robert Audi—who defends a causal account—does a nice job of bringing together the best insights of both ways of approaching the question, especially in his suggestively titled essay, “Wants and Intentions in the Explanation of Action.”

Throughout his work in action theory, Audi, in concert with the dominant causal account of reasons, emphasizes the role that desires (or, as he prefers, “wants”) and beliefs play in acting for a reason. It is hard to deny that this is as it should be; nor do I think we should want to. Clearly both do play a prominent role in our actions. Audi suggests that we can capture this want-belief pair essential to reasons by means of infinitives: “Jackie attended the funeral to support her friend.” Here Jackie wanted to support her friend and believed attending the funeral would fulfill that want.\textsuperscript{68} Audi describes the view that emerges as “a special marriage of the conative and the cognitive”\textsuperscript{69} and understands his view to be “a bit closer to the causalist theory than to the hermeneutic position.”\textsuperscript{70}

At the same time, Audi holds that the “specific causal powers” of the wants and beliefs “are partly determined by their content, and that is the very factor which, on the hermeneutic view, enables them to make action intelligible.”\textsuperscript{71} Here the causal operation of the beliefs and wants proceeds by virtue of the semantic properties of their content, and thus Audi makes his stance clear: “I do not construe wants or beliefs as events or as Aristotelian efficient causes of any kind… I may be regarded as the source of my actions and may truly say that I cause them. But surely this does not require postulating agent-causation in any sense incompatible with the explainability of the actions by appeal to my wants and


\textsuperscript{69} Audi, “Overview,” 9.

\textsuperscript{70} Audi, “Wants and Intentions,” 124. If that is true—and I wonder how many causal theorists would agree—it is nonetheless the case that Audi insists on a very important role for a “hermeneutic” account.

\textsuperscript{71} Audi, “Overview,” 15.
beliefs.” Audi’s causal account of reasons insists that the causality of the reasons cannot be separated from the hermeneutic activity of the agent herself, and action for a reason “is a discriminative response to one or more reasons of the agent.” Indeed, Audi even gives a cautious reformulation of the “logical relation” argument that was a standard argument made in defense of a hermeneutic view. After defining “reasons proper” as some sort of abstract entities that are the contents of wants and beliefs, and “reason states” as the psychological states like wants that an agent is in, Audi gives the argument:

Since acting for a reason is in some sense a causal notion, and abstract entities are not terms in causal relations, reasons proper are not causes. Thus, actions for reasons, which are in some causal way responses to reasons—the ‘for’ surely means roughly ‘on account of’—must be responses to reason states, which unlike abstract contents, can be causal factors. But an adequate account of action for a reason should also bring out the role played in such action by reasons proper. That role is in part to provide the kind of intelligibility just described in reconciling nomic and hermeneutic approaches.

For Audi then, the hermeneutic element of reasons is something that cannot be dispatched, for in doing so one would cut out the very element through which the causality of reasons operates.

Audi shows how hermeneutic and causal accounts may be reconciled, at least in part. What he does not show (at least in the tiny bit of his voluminous writings that I have read) is how a materialist could appropriate the hermeneutic side of his reconciliation (no doubt

---

72 Audi, “Wants and Intentions,” 139-140; emphasis in original. Audi’s account is clearly compatibilist.


74 Audi, “Overview,” 16. Another qualified defense of the argument may be found in Audi, “Wants and Intentions,” 141. The argument is discussed in Donnellan, “Reasons and Causes.” The nub of the argument is captured by Davidson (who rebuts the argument) in “Actions, Reasons, and Causes,” 13-14: “Since a reason makes an action intelligible by redescribing it, we do not have two events, but only one under different descriptions. Causal relations, however, demand distinct events.”
partly because he is not a materialist), though he does claim that his account “is consistent with a wide variety of philosophical and psychological accounts of the underlying dynamics.”

How wide this variety is may be open to question, and the devil will be in the details (as Dretske amply shows). But the crucial thing here is that, following Audi, one need not take causal accounts and hermeneutic accounts of reasons to be irreconcilably opposed. In the debate about reasons and causes that pitted “rationalists” and “empiricists,” at least a partial reconciliation is possible. Nevertheless, while Audi’s reconciliation proceeds in terms of beliefs and wants, in accord with a causal account, “this is not to say that hermeneutic explanations—which I prefer to call intentionalistic explanations—are dispensable or reducible.” Indeed, Audi’s account reemphasizes that the hermeneutic element cannot be eradicated from an account of reasons, but is that through which the beliefs and wants operate in their causal efficacy. An account of reasons needs to give some role to the agent’s understanding of the prescriptive address expressed in reasons if it is to do more than dismiss reasons in favor of a causal account akin to Jim’s knee jerking upward.

As opposed to causes, then, reasons involve support relations that come to expression in a prescriptivity understood by the agent as rightly directing her as to what she should believe or do. She recognizes what is called for in a way that it is possible for her to act on. In a limited sense this can, of course, be operative in instrumental reasons. Given a want that an agent has, she may see a support relation prescribing a particular course of action that rightly directs her toward what should be done, given her situation. If Jenny

75 Audi, “Wants and Intentions,” 141.

76 Audi, “Wants and Intentions,” 143.
wants to move up the ladder, she should take the transfer to Cleveland. If Tom wants to encourage his Mom, he should visit her in the hospital rather than going bowling. There is a prescriptivity operative through an appeal to the agent’s understanding, such that she can see what should be done to achieve what she wants to achieve. And, if a person has a reason to act in these situations, the causal efficacy passes through the person’s understanding of the situation. Insofar as instrumental reasons are reasons and not mere causes, they do not bypass the person (remember Jim’s leg?) but operate precisely by appealing to the person as such in her recognition of what is uniquely supported or called for in the situation, given her wants.

What is of particular interest, however, is that one may be motivated to act out of accord with one’s reasons, even when one sees the reasons. This does not make much sense in instrumental reasons (I don’t want to do what I want to do?), except as there is some moral opposition to this want in which I see that I should want something that I don’t want. More interestingly, while wanting not to do what I see I have reason to do would be strange in the case of Jane’s geometry problem (geometrical *akrasia*?), which is not a case of instrumental reason, it is quite familiar when dealing with morality. I recognize what I should do or ought to do or must do, even if I find myself utterly disinclined to do it. The prescriptivity that I come to understand has the capacity to stand over against what I want to do; it is authoritative, and it is operative through my understanding of the claim that comes to expression in the moral reason. To reiterate what was said above, the reasons I shall be concerned with in this dissertation and that the problem of moral reasons arbitrariness refers to are these sorts of reasons, namely, expressions of independent,
authoritative prescriptivity and thereby make a claim upon an agent that she is capable of recognizing and responding to in a moral situation—and which she sees as having a rightful claim over her, a claim that stands even if all her wants drive her in the opposite direction. I call these “external moral reasons.”

In the sense of reasons as I understand them, nothing that does not have the requisite prescriptivity that is open to recognition by an agent is a reason; in this sense there may be motives that are not reasons. Causes, on the order of Jim’s leg jerking upward when the doctor taps it with a mallet, are of a different order than prescriptive reasons. In such cases, when asked for an explanation of why someone behaved in a particular way, we might use this reasons talk as part of our explanation—just as we do when asked to give an explanation for why Jim’s leg jerked—but such talk arises in our explanation, and not on the side of the events in question. Such causal relations are not of interest here, and if reasons amounted to nothing more than that, then I would hold out very little hope of a solution to the problem of moral reason arbitrariness that I am laboring to make plain in the present chapter.

Additionally, I do not need to take a stand that reasons are not causes, so long as the hermeneutic element is retained as essential. Audi shows how this may be done, and even though I am not entirely comfortable with all aspects of Audi’s account, a causal account like his that runs the causality through the agent’s recognition of the prescriptivity would be sufficient to avoid moral reasons arbitrariness—so long as the moral philosophy in which it is employed can give an account of the independence and authority of this prescriptivity. Reasons that are shorn of this independent, authoritative prescriptivity are no longer reasons
in a sense obviously strong enough to solve the problem of moral reasons arbitrariness. Moral reasons arbitrariness threatens any moral philosophy with an anemic account of moral reasons.

I shall not be arguing that there actually are moral reasons that express authoritative prescriptions in some way agents can recognize. Rather, I am involved in laying out what a reason is and formulating problem of moral reasons arbitrariness, a problem I will try to show at work in the moral philosophies of Allan Gibbard, Christine Korsgaard, and John Post. Each of these moral philosophies fail to ground external moral reasons, and each of them succumbs to moral reasons arbitrariness. In this dissertation I do not need to establish that there is a moral theory that does ground such moral reasons (though I believe there is). I need only to show what the problem of moral reasons arbitrariness is, and illustrate the problem by giving some examples.

Before beginning to lay out the problem of moral reasons arbitrariness specifically, there is one—rather technical—issue about reasons that needs to be canvassed. Philosophers have had much to say about internalism and externalism in connection with moral philosophy, and it is important to place what is being advanced here within the topography of these discussions. Doing this will complete our discussion of moral reasons and orient the discussion of what moral reasons arbitrariness is. The first point of reference will be a widely read essay by William Frankena.
C. Internalism and Externalism: Sketching the Topography of Reasons

1. The Topography Sketched. In an influential essay entitled “Obligation and Motivation in Recent Moral Philosophy,” William Frankena identifies himself as an “externalist.” This language, the language of “externalism” and “internalism,” has metastasized into a philosophical argot with a life of its own since Frankena’s day, and it is not always a simple task to sort out what a particular philosopher means when using the language. In the current section, I will attempt to sketch the lay of the land with these issues and to (1) clarify what I shall mean by the terminology, and (2) briefly situate my arguments within the contemporary scene.

Perhaps the best way into these issues is to follow along with Frankena for a little while before turning to more recent formulations. Frankena lays out eight assertions that externalists object to, the first four of which must be denied by the externalist, while the latter four admit of dispute. For the moment, let’s consider only the first four theses, the ones the externalist must reject, according to Frankena:

1. That the state of having an obligation includes or is identical with that of being motivated in a certain way.
2. That the statement, “I have an obligation to do B,” means or logically entails the statement, “I have, actually or potentially, some motivation for doing B.
3. That the reasons that justify a judgment of obligation include or are identical with the reasons that prove the existence of motivation to act accordingly.
4. That the reasons that justify a judgment of obligation include or are identical with those that bring about the existence of motivation to act accordingly.\(^{77}\)

---

\(^{77}\) William Frankena, “Obligation and Motivation in Recent Moral Philosophy,” in Essays On Moral Philosophy, ed. A.I. Meldin (Seattle: University of Washington Press, 1958), 60; Frankena’s emphasis. This essay is reprinted in Thomas L. Carson and Paul K. Moser, eds., Morality and the Good Life (Oxford: Oxford University Press, 1997). I will postpone consideration of the other four assertions for the moment. They will be considered below.
If any of 1-4 can be shown to be true, Frankena argues, then externalism is false. The first thing to note when considering these four assertions is that the externalism/internalism distinction is closely tied to the justifying reasons/motivating reasons distinction. Indeed, one feels a temptation here simply to equate internal reasons and motivating reasons, on the one hand, and external reasons and justifying reasons, on the other—and Frankena, it seems to me, has perhaps succumbed to this temptation. Nevertheless, it is important to keep these categories of reasons distinct if one would not beg some significant questions in moral philosophy, for some have wondered: Could there, perhaps, be internal reasons that in some manner justify our choices?

Clearly, however, internal reasons are tied to motivation. In each of Frankena’s four assertions that must be denied by externalists, we see that obligation would be somehow logically bound to motivation or else causally bound to motivation, such that the state of being obligated “is identical” with being motivated, or “logically entails” being motivated to undertake an action, or that the same reasons that justify an obligation are the very same ones that “prove the existence of motivation” or else “bring about the existence of motivation” to act in such and such a way. But the statements above are ambiguous, for at

78 Throughout “Obligation and Motivation” Frankena appears to run internalism together with motivating reasons and externalism together with justifying reasons. I believe that the intuition probably lying behind this is at least partially correct, but simply to define matters in this way forecloses matters that must be investigated

79 In particular, we need to hold open the possibility that internal reasons may possibly have justificatory force when understood properly. Indeed, I take it that vindicating internal justifying reasons is a lynchpin of Korsgaard’s project of vindicating justifying reasons within a situation in which we human beings are the ones who “confer value on the objects of our rational choices” (Christine Korsgaard, Creating the Kingdom of Ends (Cambridge: Cambridge University Press, 1996), ix; Korsgaard’s emphasis). On the other side of the issue, it could be that external reasons might intrinsically have the capacity to motivate action. These would be the “objective values” or “intrinsically prescriptive entities” that J.L. Mackie seeks to flay in his “argument from Queerness” (Mackie, Ethics, 40)
least (1) and (2) leave out an important consideration. This has to do with what we might think of as a matter of logical priority or the direction of the influence between justification and motivation. Take (1) for instance. Perhaps the identity of the state of having an obligation and the state of being motivated to fulfill that obligation arises not from the identity relation being dominated by motivating reasons, with the justificatory force flowing from the motives, but the converse. To take an example, perhaps we are dealing with something like the Platonic Good that bears its motivational force in itself, so to speak, such that to rationally apprehend it is to thereby be motivated to pursue it. Here the justifying apprehension of the reason precedes and accounts for the motivational force of the reason. The reason motivates because it justifies, and if the justifying reason were to be undercut, the motivating force would dissipate. Perhaps then it would be helpful to think of a dimension of internalism and externalism as being plotted on a continuum between two poles—one where motivation is completely in ascendancy, with justification having no independent ground, and one where justification reigns, with motivation following along behind it. At the first pole we might find the kinds of positions that are at least attributed to Hume, while at the other pole we might views that are often brusquely dismissed as a “Platonism” of the sort that J.L. Mackie finds “queer.”

The leading idea of internalism, then, is that a reason that gives the person who sees it no motivation to act is not a reason at all. Internalism is thus “the view that the standards

---

80 In other words, Frankena has not fully allowed for a way in which an externalist might be willing to grant that “the state of having an obligation includes or is identical with being motivated in a certain way.” In certain cases, it may be that an external reason motivates simply by being intellectually understood, as with Plato’s “Good.”
of rationality binding on a person must be motivationally cogent for that person.” The externalist denies these things, arguing instead that a person can have reasons that obligate him to undertake a particular action that nonetheless leave him motivationally cold, perhaps because of a psychological defect or the deleterious effects accumulated immoral choices have had on his character (what St. Paul spoke of as one’s conscience having “been seared, as with a hot iron”). To return to the example of Ted deciding whether or not to leave a note on the car he damaged, perhaps he follows his lesser motives and takes off without leaving a note, leaving the car’s owner to fix the damage herself. But for all of that, Ted may well have a justifying reason that should rightly have trumped the motivating reasons he in fact chose to follow. Thus a basic issue between internalism and externalism within moral philosophy: Are there reasons for moral judgments or actions that somehow transcend motivations and cut across people’s subjective motivational sets, such that even if a person is not motivated to do something, he can still have a reason to do it?

At this point, our distinction between internalism and externalism is still fairly blunt. To introduce some important nuances, let’s return to Frankena’s assertions about the differences between internalism and externalism, considering assertions 5-8 that Frankena puts forward. According to Frankena, the externalist will likely deny

---


82 I Timothy 4:2. Many commentators argue that Paul did not write I Timothy, but getting into that issue obviously is of no concern here.
5. that, although justifying a moral judgment does not include giving exciting [i.e., motivating] reasons for acting on it, it presupposes the existence, at least potentially, of such excitement.

6. that saying or being said to have an obligation presupposes one’s having motives for doing the action in question

7. that assenting to an obligation entails feeling or having a disposition to feel at least some inclination to act in the way prescribed

8. that one can know or “see” or think that one has a certain obligation only when one is in a favorable conative state with respect to performing the act in question.  

These four assertions may be admitted by some externalists, according to Frankena, although they “would be denied by a really ‘compleat’ externalist.” In each case, Frankena leans toward rejecting the assertion, though he seeks to show how one could still be a card-carrying externalist and accept 5-8 (and he gives some examples of some who do—for example, Scheler and Hartmann maintain (8) or something similar).

It’s helpful here to compare Frankena’s assertions with the taxonomy of various internalist and externalist positions as presented by Russ Shafer-Landau in his *Moral Realism: A Defense*. In what follows I will accept major elements of Shafer-Landau’s taxonomy and try to draw connections between them and Frankena’s internalist assertions. I will adopt Shafer-Landau’s terminology.

At the heart of Shafer-Landau’s taxonomy is a distinction between judgment internalism/externalism and reasons internalism/externalism. Reasons externalism holds that “our reasons can exceed our capacity to appreciate them, or to be motivated by them,”

---

83 Frankena, “Obligation and Motivation,” 60; Frankena’s emphasis.

84 Frankena, “Obligation and Motivation,” 60.
while reasons internalism holds that reasons cannot thus outstrip our motivations.  

Questions of judgment internalism/externalism, on the other hand, have to do with how we understand someone’s sincere judgment that an action is right or wrong. “Judgement internalist theories specify necessary connections between moral judgements, on the one hand, and either motives or justifying reasons, on the other.”86 This fundamental distinction partially reflects the difference Frankena sees between assertions 1-4 and assertions 5-8. The former assertions, which address a more fundamental issue, have to do with questions of whether justifying reasons are somehow logically or closely causally tied to motivating reasons and dependent upon the force of a person’s motives. The logical and even ontological status of justifying and motivating reasons is at stake in assertions 1-4. One’s stance with respect to internalist assertions 1-4 (suitably qualified, as above) places one on

85 Shafer-Landau, *Moral Realism*, 189. Jean Hampton talks of this same distinction as being between “Motivational Internalism” and “Motivational Externalism.” See also Hampton, *The Authority of Reason*, 58. I will follow Shafer-Landau’s usage.

Shafer-Landau doesn’t deal extensively with the question of “ideal motives,” working instead with the motives we actually have. He does, however, realize that we do bear at least some responsibility for our actual motives and that we can choose to cultivate better, perhaps more “ideal” motives. The key question, of course, will be specifying this ideal. Materialism may somewhat complicate the attempt to do so. In any case, to specify how I would be motivated if I were ideally motivated and then specifying the relation of those motives to a person’s actual motives would open quite a number of difficult philosophical issues for the person inclined toward such a view. My view is that a philosopher willing to countenance a much richer metaphysic than materialism could make some headway on these issues. See, for example, Linda Trinkaus Zagzebski, *Divine Motivation Theory* (Cambridge: Cambridge University Press, 2004).

I am skeptical that the Spartan metaphysics of materialism can get very far with this, however. Zagzebski’s version of the theory places a great deal of weight on an “exemplar,” and the materialist’s lack of a Divine Person whose motives are in fact ideal changes a great deal about how such a theory could unfold. Tom Carson’s not wholly unrelated “Divine Preference Theory” also prominently features the preferences of a Divine Person. Carson is concerned to preserve a “fallback” notion of rationality that will serve in place of the preferences of a Divine Person, should it turn out that no such Person exists. See Carson’s sketch of an “informed-preference theory of rationality” (Carson, *Value and the Good Life*, 230-39) for some thoughts about how this might work. Dealing with this issue in the depth it would deserve would require another chapter, probably focused on the work of Richard Brandt. Here I must rest content to say that I am less sanguine about the prospects of this tack than Carson is. I have registered some of my hesitations in Seeman, “Whose Rationality? Which Cognitive Psychotherapy?”

86 Shafer-Landau, *Moral Realism*, 142-3; Shafer-Landau’s emphasis.
the map of being either a reasons internalist or a reasons externalist. To accept reasons internalism is to accept that we can have no reasons that do not motivate us to action (though through accidie or akrasia we may yet fail to do the action), while reasons externalism insists upon reasons that transcend a person’s “subjective motivational set.”

Assertions 5-8 pass into different territory, and they partially match up with Shafer-Landau’s distinction between judgment internalism and judgment externalism. In particular, (7) clearly is a statement of this distinction. In (7), what is at issue is assenting to an obligation, which is what Shafer-Landau captures with his talk of “making a sincere judgment” that something is right or wrong. When someone does actually assent in this way, the judgment internalist would hold that this person must be motivated to perform the action—perhaps with the exception of her being in a state of accidie or akrasia. 87 I believe (6) also relates to the judgment internalism/externalism distinction, although this is not quite as clear as with (7). Certainly, if the issue is whether, when a person sincerely says of herself that she has a moral obligation (from within her own first person standpoint), she must in fact be motivated to act in accordance with the judgment she is saying she holds to, then we are wondering about questions of judgment internalism/externalism. The confusion comes when we consider Frankena’s introduction of a second or third person standpoint, in which I look at someone else and say that she has an obligation to take a particular course of action. Clearly for that person herself and the statement she has made it is right to apply the judgment internalism/externalism distinction; but it’s not clear that her

87 Shafer-Landau formulates “weak” and “strong” forms of judgment internalism, depending on whether or not they make allowance for akrasia. When Michael Smith defends a form of internalism, he explicitly rejects a strong form of judgment internalism, because it “commits us to denying that, for example, weakness of the will and the like may defeat an agent’s moral motivations while leaving her appreciation of her moral reasons intact.” See Michael Smith, The Moral Problem (Oxford: Blackwell, 1994), 61.
“being said to have an obligation” by someone else brings up the same issue for her, since the judgment is not her own. It is being attributed to her from without, and it’s not clear that someone can know from this external vantage whether she is sincerely making a judgment. I conclude that (6) either collapses into (7) and thus falls under the judgment internalism/externalism distinction, or else that (6) is making a point that cannot be sustained, because it presumes we can know from the outside whether another person sincerely makes a particular judgment.

What are we to make of (5) and (8) then? Let me not belabor these points here, for they are not central to what follows. But so that the taxonomic question can have some level of completion (at least with respect to Frankena’s assertions), I will say a few words. First, I am far from sure that (5) does not belong with assertions 1-4 as an assertion of a form of reasons internalism. The question would turn on how we are to understand the statement that justifying a moral judgment “presupposes the existence” of motivating reasons. This seems to be an instance of the unconditional priority of motivating reason over justifying reasons, and it would thus fall under the distinction between reasons internalism and reasons externalism.

Turning lastly to (8), this is an epistemological question of the conditions one needs to be in in order to know that one has a justifying reason to perform an action. Some conative states may lead one to miss the fact that one has a justifying reason to perform an action. But this epistemological issue does not address the issue of whether justifying reasons exist independently of motivating reasons, or the issue of whether making a sincere moral judgment necessarily motivates one to act upon that judgment. If this is correct, then (8),
while an interesting matter, does not pertain directly to distinctions between internalism and externalism.

All this seems to leave us with two major distinctions in matters of internalism and externalism in moral philosophy. First, there is the distinction between reasons internalism and reasons externalism. To repeat, this distinction deals with whether a person can only have a justifying reason for an action if he also has a motivating reason moving him to perform the action, where the motivation does not arise from the justifying reason (as with Plato’s “Good,” for example). Second, the distinction between judgment internalism and judgment externalism has to do with whether a person who makes a moral judgment about a particular action necessarily has a motivating or justifying reason for that action.

Judgment internalism comes in three forms, according to Shafer Landau’s taxonomy: Motivational judgment internalism, reasons judgment internalism, and hybrid judgment internalism. Motivational judgment internalism holds that the person who makes a sincere moral judgment has a motivating reason to do the action in question. Shafer-Landau states that this form of internalism “has itself lent important support to a variety of antirealist positions in metaethics.” Reasons judgment internalism, on the other

---


89 Shafer-Landau, *Moral Realism*, 143. The support for antirealist positions runs as follows: If all moral judgments must motivate to some degree, then they cannot (under a Humean theory of motivation) be beliefs, but must express desires. Only in this way could motives and judgments be thus inseparable; they must be expressions of desires and thus inherently motivating (this argument is summarized in Shafer-Landau, *Moral Realism*, 4, 120-21). I find this very interesting, for I am attracted to a chastened moral judgment internalism, while also defending a realist metaethical stance.

One obvious way of challenging this noncognitivist argument is to challenge the Humean theory of motivation. One might also point out that judging that one has a reason is different than one’s actually having a reason. Sincere judgment does imply an endorsement or acceptance that cannot readily be understood in a way that avoids talking of the agent’s motivation, and yet for all of that one’s sincere judgments could miss reasons that really are there. This is, of course, quite inhospitable to noncognitivism. Of course I am not meaning to resolve such issues in this one footnote.
hand, is the notion that a sincere moral judgment necessarily gives a justifying reason for an action. This position seems so wildly implausible to me, that I have a difficult time stating exactly what could be meant by it. If I understand the position correctly, one would be saying that if one makes a sincere moral judgment then one is thereby supplied with a justifying reason for that action. But surely many sincere moral judgments have been wrong. Perhaps some Nazis sincerely judged that a life of protecting blood and soil from Jewish corruption was morally right, but it is not plausible to say that they were thereby provided with a justifying reason on which they were right to act. That whatever one is motivated to do is—by that very fact—justified; surely that is a bitter pill to swallow.

Lastly, hybrid judgment internalism “states that a person sincerely judging an action right has reason to be motivated to perform that action.”

Shafer-Landau states that what he calls “hybrid judgment internalism” is the same position Michael Smith calls “the practicality requirement” in Chapter 3 of *The Moral Problem*. Smith’s formulation runs as follows: “If an agent judges that it is right for her to ϕ in circumstances C, then either she is motivated to ϕ in C or she is practically irrational.” It seems to me (with some hesitation) that Shafer-Landau is mistaken about how his terminology matches up with Smith’s terminology. Specifically, the “practicality requirement” seems to be what Shafer-Landau calls the “weak” form of “Motivational Judgment Internalism” in which “a person sincerely judging an action right is motivated to some extent to comply with her judgments.”

---


form of motivational judgment internalism is “weak” because it makes allowance for “akrasia” and other motivational or constitutional defects. Indeed, it is difficult to see exactly what difference Shafer-Landau hopes to capture by distinguishing between weak motivational judgment internalism and hybrid motivational internalism. The difference between them is that the latter states that the agent has reasons to be motivated to \( \phi \), while the former states that the agent is motivated to act on the judgment, but in a way that may be overturned by akrasia or the like. It’s difficult to see how the agent is motivated to \( \phi \) in the absence of reasons for doing so (which itself points to an ambiguity: when Shafer-Landau speaks of “reasons” to \( \phi \), does he mean motivating reasons or justifying reasons?). Additionally, the qualification on weak motivational judgment internalism that restricts the motivation to instances where it is not overcome by akrasia seems to account for the purpose behind saying that the agent merely “has reasons” to be motivated to \( \phi \) rather than saying that the agent is motivated to \( \phi \). Presumably the purpose is to create a gap between being motivated to \( \phi \) and actually \( \phi \)-ing. But the allowance for akrasia already accomplishes this.

In any case, it is clear that sorting out this mare’s nest of internalism/externalism terminology and the different ways it is employed by different philosophers is not easily done, especially when things get sliced as thinly as they are in Shafer-Landau’s taxonomy. That said, some basic distinctions are clear, and these distinctions will be enough for the purposes that will be pursued through the rest of the dissertation. In particular, we have seen the major distinction that runs down the middle of William Frankena’s internalist assertions. It is fair, I believe, to think of this fundamental internalist/externalist divide as
an ontological divide: Are justifying reasons in some way or another reducible to motivating reasons, so that the former do not have an independent existence of their own? Shafer-Landau calls this the question of reasons internalism/externalism, and I will follow him in this. It is quite important to keep this ontological question separate from the question of the relation between an agent’s sincere moral judgments and that agent’s motivations. The various positions with respect to issues of sincere moral judgment and motivation I will label issues of judgment internalism/externalism.93

2. Standing within the Topography of Reasons. I do not believe that reasons internalism has the resources to avoid the problem of moral reasons arbitrariness (though I do not argue here that there could not possibly be such a theory), and that if we are to avoid this “coal pit” we should search earnestly for a ground for external moral reasons. As I have said, this is an ontological question, and thus I find myself diametrically opposed to statements like that of Thomas Scanlon. Faced with the worry that moral judgments may admit of no right or wrong answers, Scanlon admits that “it may seem that in order to answer it we must provide a metaphysical account of the subject matter of judgments of this kind. I believe that this is not what is necessary, however. The question at issue is not a metaphysical one.”94 Certainly many—if not most—prominent moral philosophers writing today would cast their lot with Scanlon on this. If we can locate the proper procedure, the

93 As I am using the term “judgment” here, I intend the term very broadly. “Endorsement” or “acceptance” would count as “judgments” in my sense of this term. The intention here is to make it clear that judgment internalism/externalism is neutral with respect to cognitivist and noncognitivist views.

94 Scanlon, What We Owe, 2.
proper way of reasoning, the problem will be solved; and if we cannot solve the problem in this manner, then we will have to make do with some stand-in for independent, authoritative prescriptive moral reasons. As Scanlon puts it, if this question of the procedures or methods of reasoning about moral judgments could be solved, then “no interesting question would remain about the ontology of morals.” As far as I can tell, this is just mistaken. Rather, if external moral reasons are ontologically indefensible, it is not clear to me what hope we have for avoiding moral reasons arbitrariness. Indeed, the dissertation will argue that three materialist attempts to ground moral philosophy fall into moral reasons arbitrariness, and the culprit appears to be their inability to provide us with external moral reasons. If I am right, moral philosophy needs a different way forward than what is found in Gibbard, Korsgaard, and Post (Millikan), a way that grounds external moral reasons that express authoritative prescriptions that are not beholden to the subjective motivational set of any contingently personal being and uniquely justify some courses of action over others.

In matters of reasons internalism/externalism then, I locate myself on the reasons externalist side of the divide. This reasons externalism, however, I would combine with a form of judgment internalism. William Frankena rightly notes that some philosophers (whom I am calling reasons externalists) that reject internalist assertions 1-4, may nevertheless find one or more of the internalist assertions 5-8 plausible. I would count myself among these. Specifically, I find (6) and especially (7) to be very attractive, and holding to these positions would make one a judgment internalist, according to Shafer-Landau’s taxonomy. But such

---

95 Scanlon, *What We Owe*, 2.
Judgment internalism is fully consistent with being a reasons externalist. As Frankena observes, holding to (7) “may only mean that we commonly believe that all men are psychologically so constituted as to be moved by the recognition that something is right. It need not mean that this is logically necessary… [O]ne may hold rightness to be an external characteristic and yet claim that we are so made as necessarily (causal) to take an interest in it.”96 This is precisely how I would characterize the sort of judgment internalism I hold to. Once someone does in fact make a sincere moral judgment, that person does have a motivating reason to act in accord with that judgment. This strikes me as a matter of psychological fact; as such, however, it is also defeasible. Weakness of the will is also a matter of psychological fact, as is the attractiveness of the motivating reasons that can sometimes array themselves quite impressively against the force of a sincere moral judgment. Sometimes I do not do what I (know I should) want to do.

Reasons externalism can sit quite comfortably beside judgment internalism of the sort I find attractive; moreover, there is no need to take the latter as an ally of noncognitivism.97 In the pages that follow, not much will be said about judgment internalism. Reasons externalism, on the other hand, will be important in what follows.

In calling moral reasons external, then, the point (at least in these pages) is that they are independent of the subjective motivational set of any contingent personal being, that the reason Jane has for employing the Pythagorean Theorem and Ted a reason to leave the note irrespective of the contents of their subjective motivational sets. The relevant judgments or

96 Frankena, “Obligation and Motivation,” 65; emphasis in original.

97 See note 89, above.
actions are uniquely called for independently of their coming to see them or being motivated in relation to them.

With this reasons internalism/externalism distinction in place, the account of reasons being developed here is complete and we can begin to make plain the problem of moral reasons arbitrariness. Again, it should be made clear that it is not the point of this dissertation to argue that we do in fact have external moral reasons, as that notion has been developed here. But subsequent chapters will argue that the materialist moral philosophies of Gibbard, Korsgaard, and Post fail to provide external moral reasons, and that this leaves them with no clear way to avoid moral reasons arbitrariness. The nature of this problem will begin to emerge as we begin by looking at a materialist philosopher who more or less embraces moral reasons arbitrariness as an outworking of his materialism (“Science,” as he calls it), namely, Gilbert Harman. A clear-eyed, flinty Scientist sees through the surface of morality to what is really going on. But what Harman claims to embrace, keeps another materialist philosopher, Christine Korsgaard up at night. Her question, asked as a justificatory question, is “Why be moral?” if you’ve seen through your moral reasons to non-moral forces that lead them around by a hook in the nose. The issue between Korsgaard and Harman will prove to be a helpful avenue of approach to seeing the problem of moral reasons arbitrariness.
II. Korsgaard and Harman, Dictators and Criminals: An Initial Approach to the Problem of Moral Reasons Arbitrariness

A. Korsgaard’s Worry

At the start of her Tanner lectures at Cambridge University, Christine Korsgaard gives voice to a serious concern: How are we to understand the justification of our moral actions? As she notes, “the day will come, for most of us, when what morality commands is hard … And then the question—why?—will press, and rightly so. Why should I be moral?” Korsgaard makes it clear that she is not after a merely explanatory account, for such an account may leave us thinking that we “see through” the commands of morality to what is “really” going on (I am feeling the force of my evolutionary history or my culture or the like). Morality in that case would be “explained,” but not justified. Its authority would be undermined. It loses its grip on the moral agent qua moral agent. “A theory that could explain why someone does the right thing—in a way that is adequate from a third-person perspective—could nevertheless fail to justify the action from the agent’s own, first-person

---

98 Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), 9; Korsgaard’s emphasis. Korsgaard wants her listeners to be very clear what this question is that she is asking. To ask the question as she asks it is not, she insists, a call for an explanation of our motivations for moral behavior. Instead it is to “request an account of rightness” (9). She continues: “we are not looking merely for an explanation of moral practices. We are asking what justifies the claims that morality makes on us” (9-10). Korsgaard calls this “the normative question” (10).

99 Korsgaard couches this worry in terms of an “evolutionary theory,” and poses the following question: “ask yourself whether, if you believed this theory, it would be adequate from your own point of view” (Korsgaard, *Sources*, 14). Korsgaard connects this worry to Bernard Williams’ reflections on what he calls “transparency.” Here’s how Korsgaard puts it:

If a theory’s explanation of how morality motivates us essentially depends on the fact that the source or nature of our motives is concealed from us, or that we often act blindly or from habit, then it lacks transparency. The true nature of moral motives must be concealed from the agent’s point of view if those motives are to be efficacious (Korsgaard, *Sources*, 17)
Korsgaard’s worry hinges on a distinction between “motivation” and “justification” in morality, or, as we might put it, between what functions as a motivating reason for our behavior and a justifying reason for our action. In his classic essay, “Obligation and Motivation in Recent Moral Philosophy,” William Frankena draws this distinction with characteristic clarity: “we should distinguish between reasons for acting and reasons for regarding an action as right or justified… it is not plausible to identify motives with reasons for regarding an action as morally right or obligatory. At any rate there is a prima facie distinction to be made between two senses of ‘ought’ and two kinds of reasons.”

What I am motivated to do often conflicts with what I take myself to be obligated to do, or to be the right thing to do. What I cite as having motivated my action may do little or nothing to justify that action. For example, I may cheat on an exam because I believe it will help me get a better grade, and thus a better job and more money. Such motives are perfectly comprehensible, but they do not justify my choice, even though we clearly understand how

---

100 Korsgaard, Sources, 14

they motivate my choice. The justification I do or do not have in my actions stands over against my motives, and in some sense judges whether it would be right to act on the basis of one motivating reason or another.

Frankena thus identifies an ambiguity lurking in talk about reasons: we often fail to distinguish whether we are talking about motivating reasons or justifying reasons (where only the latter are characterized by authoritative prescriptivity). “When A asks, ‘Why should I give Smith a ride?’ B may give answers of two different kinds. He may say, ‘Because you promised to,’ or he may say, ‘Because, if you do, he will remember you in his will.’ In the first case he offers a justification of the action, in the second a motive for doing it. In other words, A’s ‘Why should I…?’ and ‘Why ought I…?’ are ambiguous questions.”

In one sense of “should” or “ought,” people act in accord with their motivations, and their actions simply dovetail with the structure of their “subjective motivational set.” But in another sense—and this is the sense of Korsgaard’s worry and the concern of this dissertation—there is an “ought” or “should” that stands over against a person’s motives. This is the place where “doing the right thing” can be very costly. Think here of the holocaust rescuers as opposed to those who were complicit with the Nazis or simply kept their heads down as well as they could. The rescuers did what they should have done or ought to have done in this costly sense; they responded to authoritatively prescriptive moral reasons in the face of

103 The phrase is Bernard Williams’. See Williams, “Internal and External Reasons.” The contents of a SMS can be desires, ends, aims, patterns and dispositions of evaluation and emotional reactions, and various loyalties and projects, and these may change in response to deliberation and need not be egoistic in all cases (103). Reasons tied to a person’s SMS are that person’s “internal reasons,” as Williams calls them. Williams argues that there are no “external reasons,” but only different ways of talking about a person’s (internal) reasons. I think his argument helps itself to resources that the internal reasons theorist cannot use without begging the question, but I do not need to argue that point here.
very powerful motivating reasons not to, doing what was *uniquely justified* by those reasons.\(^{104}\) When doing the right thing gets tough, where do obligations find grounds that stand against countervailing motives?

The opposition of justifying reasons to motivating reasons is a common feature of moral experience that highlights the importance of keeping these two types of reasons distinct. One’s motives—what one would “like” to do—can be very clear, and yet one may hold back, convinced that to do so would be to violate a duty, or perhaps searching for the right thing to do in the situation. As Frankena emphasizes,

When a man thinks that something is his duty, what he thinks is that it is a duty independently of his thinking so (and independently of his wanting to do it); and when he asks what his duty is, he implies that he has a duty that he does not yet recognize, and what he is seeking to know, as it were, is what it would have been his duty to do even if he had not discovered it. Thus there is a sense in which one has a moral obligation even if one does not recognize it as such.\(^{105}\)

People look for their duty, they feel guilty when they do not do what they ought to have done, they feel something standing over against their motivations—something that claims the right or authority to trump those motivations, even where they choose not to follow its dictates.\(^{106}\) But what justifies the claims of the “should” and the “ought” of justifying reasons? When doing what one has a justificatory reason to do costs a great deal in terms of

---


\(^{106}\) Charles M. Taylor, “Ethics and Ontology,” *Journal of Philosophy* 100, no. 6 (June 2003): puts it this way: “Ethics involves a range of ‘values’ that are essentially understood to be on a different level, to be in some way special, higher, or incommensurable with our other goals and desires” (308).
what one would otherwise be motivated to do, why should a person do it? How are we to understand the ability of justifying reasons to trump motivating reasons? Why be moral, in this justificatory sense? This is what worries Korsgaard: How can the putative authority of justifying reasons—the authority felt in people’s everyday lives (our first person perspective)—be sustained?

The worry and the distinction behind it strike me as critically important for moral philosophy. For if the distinction cannot be sustained, the reasons to which we take ourselves to be responding as moral agents are finally vehicles through which causal forces exert their sway over us. Frankena diagnoses the problem correctly: if by moral “reasons” we can only mean motivating reasons, we risk a morality that is forced “to trim obligation to the size of individual motives.” At the end of the day, I believe neither Korsgaard nor Gibbard nor Post can marshal the resources to meet this problem. Rationality, as they are able to construe it from within their materialist convictions, cannot bear the weight—or so I shall argue. But, as Korsgaard clearly sees, the consequences of a collapse of justifying reasons into motivating reasons would be worrisome indeed.

B. Dismissing the Worry: Gilbert Harman’s Critique of the Justifying Reason/Motivating Reason Distinction

Not everyone has thought that the distinction between justifying reason and motivating reasons is important or even real. Gilbert Harman in particular has critiqued the idea of a distinction between motivating and justifying reasons. Harman builds his case for moral relativism on the nature of human moral judgments. “Just as the judgement that

---

something is large is true or false only in relation to one or another comparison class, so to, I will argue, the judgement that it is wrong of someone to do something is true or false only in relation to an agreement or understanding.”

What we understand as justifying reasons end up being whatever a person is motivated to endorse given an agreement with others to act in a particular way that they have all implicitly agreed to. At the heart of this account we find what Harman calls an “inner judgment.” An inner judgment is relative to the motivations a particular person has to behave in a particular way, a way that is “moral” or “immoral” relative to a particular group that is of concern to the individual. This group is of concern to the individual because its members have the ability to aid or to hinder the interests of the individual, a capacity that is exercised frequently in all of our lives in many ways.

As Harman puts it, “the main reason why a person accepts the principles he or she does is that it is in his or her interest to do so if others do too.” And so we enter into an “implicit agreement” with one another. It is “useful” for us to do so, and there is no mystery in understanding the kinds of reasons operative in the “inner judgments” that underlie the phenomena of moral behavior: they are instrumental reasons that motivate us on the level of our interests in realizing certain ends we have. We act in “moral” ways—as those are understood by people in the relevant groups—because we need their participation.

---

108 Gilbert Harman, *Explaining Value and Other Essays in Moral Philosophy* (Oxford: Clarendon Press, 2000), 3. This book is a collection of essays, the first of which is a reprint of a well known article Harman published in 1975, entitled “Moral Relativism Defended” (pp. 3-19). In this essay Harman develops the notion of “inner judgements” in this essay, where he argues that such judgments “imply that the agent has reasons to do something that are capable of motivating the agent,” (10). In terms of the taxonomy I am following, Harman hews very closely to reasons internalism. Justificatory reasons are thus tied to the motivations of the agent.

109 Harman’s position here is quite similar to one taken by Allan Gibbard; see Chapter 3.

if we are to realize most of our aims in life, and their hostility can make things difficult indeed as we attempt to realize the ends we hope to achieve.

Reasons for moral behavior, then, are linked to our endorsement of the “moral” conventions of a group because of our interest in having the members of that group abet the ends that interest us, or at least prevent their active hindrance of those ends. Harman states it thus: “I want to argue that the source of the reasons for doing $D$ [a particular moral action] that $S$ ascribes to $A$ consists in $A$‘s sincere intention to observe a certain agreement.”¹¹¹ Reasons for moral behavior cannot be separated from a person’s intention to observe an agreement; the source of moral reasons is our acceptance or endorsement of an agreement shared by members of a social group relevant to our interests. Thus, inner judgments “have two important characteristics. First, they imply that the agent has reasons to do something. Second, the speaker in some sense endorses these reasons and supposes that the audience also endorses them.”¹¹² Moral reasons cannot be separated from endorsement, this latter being the glue that binds the individual’s actions to the conventions of society and thereby affords a scientifically respectable explanation of moral behavior.

That Harman’s explanation of moral behavior is “scientific” speaks highly in its favor, he believes. Harman does not hesitate to claim science for his way of proceeding. There are, he declares, two ways of approaching moral philosophy, and the difference “is, to put it crudely, a difference in attitude toward science.”¹¹³ The good way of doing things is

the one that insists that “we must concentrate on finding the place of value and obligation in the world of facts as revealed by science,” while the other way of doing things stubbornly insists on attending to “ethics proper.” The problem here is that “ethics proper” shares in a distasteful disregard for science, focusing on agents and actions and other scientifically undigested notions. Much better to follow science where it leads and work out an understanding of moral behavior in the terms thus provided.

The terms left to us in Harman’s approach are “inner judgment,” “agreement,” “social convention,” “endorsement,” and “reasons” (where these are traced back to interests and “motives”). And what all of this affords us in “ethics proper” is moral relativism. As Harman would have it, once we understand that “morality” is to be explained in terms of inner judgments, we see that moral reasons vary from person to person according to the relevant comparison group—the group with which we happen to enter into an “implicit agreement.” This loose contractarian element in Harman’s moral philosophy means that his moral relativism is not a subjective free-for-all, but is bound to a narrow spectrum of live options that happen to be open to a person. Thus, aliens from outer space would have no reasons to observe our moral conventions, since the conventions we have as a group are irrelevant to the comparison group within which their inner judgments are formed. They might happen to act in accord with our conventions, but they would not act because of them.

---


115 Perhaps I should say: “…not obviously a subjectivist free-for-all…”, for one does wonder about the status of moral reasons for an individual who comes to believe a view like Harman’s. If you come to believe your justificatory reasons arose from an “implicit agreement” with a group of people you happened to have been thrown into by the vicissitudes of history, and if you have no additional story to tell about some other, independent, authoritative grounds on which those moral reasons might be grounded, then perhaps the person who believes Harman’s account is cast back onto one’s own subjectivity more than Harman would like to admit. More will be said about these kinds of issues in the next section.
Or, consider two more examples: Those who scientifically explain morality (as Harman understands such an explanation, anyway), “assume that, for Hitler, there might have been no reason at all not to order the extermination of the Jews. They assume that the contented employee of Murder, Incorporated might have no reason at all not to kill Ortcutt.”

The moral reasons we have owe their existence to our endorsement of the prescribed behavior of a group which is relevant to the success of our ends. Individuals and groups irrelevant to those aims—Earthlings, Jews, or Ortcutt—may be left out of account.

Now, this is a lot to swallow, and at this point in one's reading of Harman's argument one may have had enough. It's not clear that failing to recognize the reasons that exist means that there are no such reasons, even if all contingent personal beings do not find themselves motivated to act on these reasons. Perhaps all contingent rational beings might someday find that they have reasons (in Harman’s sense) to torture dogs that do not make money in gambling on dog fighting, or that they find no reasons to prohibit the dismembering of homo sapiens who do not yet perform the functions determined requisite for being a “person.” Could they all be wrong? Or, echoing Sartre, would we have to say that these actions would then be the truth of humanity?

Can moral reasons be understood to be more robust than Harman allows? If so, how might they be grounded?

Returning now to the crucial issue of whether or not we should distinguish between justifying moral reasons and motivating moral reasons, it may be wondered whether

---


117 In terms that will be developed below, the interests and well-being of such people can viably be disregarded as one decides what to do.

Harman is trading on an ambiguity over the idea of having a “reason” to do something, eliding a critical distinction and thus effectively collapsing the discussion into terms friendly to what science (as Harman understands it) “reveals.” Harman countenances only internal reasons, but perhaps there are external reasons that are there anyway—even if you or I or indeed all contingent rational beings fail to acknowledge or endorse them. Think again of a mathematical example, such as the sense in which there is a reason of some sort for contingent rational beings to think that $2+2=4$, even if they have not yet developed an understanding of mathematics. Should some such beings, in the course of beginning to work out an understanding of mathematics not come to this conclusion—holding, say, that $2+2=5$—there is a real sense in which they would be wrong, even if they all came to the same conclusion. The same holds for a belief in the chemical composition of water. But, at least in the moral realm, Harman rejects as unscientific external reasons that authoritatively prescribe a particular action. Do such reasons exist? More importantly for the purposes of this dissertation, if they do not and only motivating reasons exist of the kind Harman will allow, does moral reasoning become infected by a problematic sort of arbitrariness?

\[119\] Internal reasons theorists typically account for these sorts of problems by an appeal to “full rationality” (an appeal that I find wholly unconvincing). One might wonder what such full rationality is getting right and why it is led in one unique direction. As I noted above, I believe Williams and others who argue that external reasons do not exist and that there exist only internal reasons beg some important questions, but that would be a different argument.

Harman simply denies that there are external reasons that prescribe actions authoritatively and by which people may come to see a particular course of action as uniquely justified. Little may be salvaged from the concerns moral philosophers have had about reasons for moral action. Harman proleptically addresses externalist objections to his critique of external justifying reasons. Taking issue with the same essay by William Frankena quoted above, Harman notes that an “externalist” might object to Harman’s own internalist arguments “by claiming that they rest on an ambiguity in the term ‘reason’, which might mean either ‘justifying reason’, that is, reason to think one ought to do something, or ‘motivating reason’.”\(^\text{121}\) As Harman summarizes the objection to his discussion of reasons, the attempt to limit reasons to matters of motivation and endorsement fails to recognize that “if a moral demand applies to someone, that person has a compelling reason to act in accordance with that demand. For here the relevant type of reason is a justifying reason: if a moral demand applies to someone, that person has a compelling reason to think he or she ought to do that thing.”\(^\text{122}\) Harman does a nice job here of anticipating the objection that I—and no doubt many others—have to his treatment of moral reasons.

Harman offers two reasons for thinking that such misgiving about his position are misplaced, to which I will add a third drawn from other things he says. First, Harman says that this is “an ad hoc manoeuvre,” with no rational support adduced in its favor.\(^\text{123}\) Second, a regress threatens in which the reasons we have for thinking we “ought” to φ are themselves justifying reasons resting on an “ought,” and so on. And, finally (and, I think,\(^\text{121}\) Harman, *Explaining Value*, 31.\(^\text{122}\) Harman, *Explaining Value*, 31.\(^\text{123}\) Harman, *Explaining Value*, 31.
most basically for Harman), authoritatively prescriptive moral reasons have no ready explanation in terms of science, as he understands it. It is far from clear how justification, normativity, reasons, actions, agents, and the whole package of “ethics proper” can be made to fit into science as understood within materialism.

Before going on to reply to Harman’s concerns, I’d like to point out that, for the purposes of my argument in this dissertation, I don’t need to. Strangely enough, though I am convinced that Harman is deeply mistaken on some crucial points, he is a sort of ally for the position I am advancing in this dissertation, an illustration of the point I hope to make with respect to the moral philosophies of Allan Gibbard, Christine Korsgaard, and John Post. Indeed, there is a sense in which I want to endorse Harman’s arguments as correctly working out the way a materialist metaphysic shapes moral philosophy. Where Harman sees materialist science undermining external reasons that authoritatively prescribe moral actions and says, in effect, that such reasons must therefore be abandoned, I agree with him about the moral ramifications of his metaphysic, but think that this gives us good grounds for questioning the metaphysic.

Harman believes that his materialist metaphysic “is no mere prejudice in favor of science; it is an inevitable consequence of intelligent thought.”¹²⁴ The person who arrives at this point of clarity and becomes a materialist will suppose “that all aspects of the world have a naturalistic location” and will insist on finding the location of things like “moral values” in the world that Science has delivered to us.¹²⁵ “My claim,” says Harman, “is that,

¹²⁴ Harman, Explaining Value, 98.
¹²⁵ Harman, Explaining Value, 84.
when one takes this attempt seriously, one will tend to become sceptical or relativistic.”

Unlike the factual beliefs that Science renders up for our consumption, moral beliefs do not need to be explained by a world of independent objects. In contrast, “our having the moral beliefs we do can be explained entirely in terms of our upbringing and psychology, without any appeal to an independent realm of values and obligations.” Patently non-moral causes lead what we human beings individually regard moral reasons wherever they will, so that our putative moral reasons are really vehicles through which these non-moral forces operate on us. In short, moral reasons are arbitrary, varying with forces completely unbeholden to morality.

This arbitrariness also emerges as we realize that in the world Science gives us our moral judgments are not uniquely justified. Instead we find—in what Harman calls a “soberly logical thesis”—that “just as the judgement that something is large is true or false only in relation to one or another comparison class, so too… the judgement that it is wrong of someone to do something is true or false only in relation to an agreement or understanding.” Thus Harman informs us that “many of us believe on reflection that different people have reasons to observe different moral requirements, depending on the moralities those people accept.” As was seen above, moral reasons are accepted in relation to an agreement that people make that serves their self-interest in some way, thus giving an account of moral reasons that is tied to motivations readily understood in

---


127 Harman, *Explaining Value*, 82.


categories amenable to Science. On coming to see this, we see that someone else’s moral reasons that stand in conflict with my own share the same basic “justificatory” structure. Neither of them is uniquely justified in a way that the other person cannot also “justify” her own, conflicting moral judgments. Again, Harman has led us into moral reasons arbitrariness.

Moreover, according to Harman, appeals to “rationality” do not change matters. Instead, “on any plausible characterization of reasonableness and unreasonableness (or rationality and irrationality) as notions that can be part of the scientific conception of the world, the absolutist’s claim is just false. Someone can be completely rational without feeling concern and respect for outsiders. But, of course, this reply appeals to naturalism.”

Indeed. Harman continues to press his point by introducing a criminal who finds that he has no reasons not to murder others. “What one morality counts as irrational or unreasonable, another does not. The criminal is not irrational or unreasonable in relation to criminal morality, but only in relation to a morality the criminal rejects. But the fact that it is irrational or unreasonable in relation to this other morality not to have concern and respect for others, does not give the criminal who rejects that morality any reason to avoid harming or injuring others.” The problem for his non-relativist opponents, as Harman sees it, is that “we must give a naturalistic account of reasons and we must give empirical grounds for supposing someone to be irrational or unreasonable.”

---

wisely follows Science at this point, is faced with an opponent who in rejecting “naturalism in favour of autonomous ethics relies on an unreduced normative characterization of rationality and irrationality (or reasonableness and unreasonableness).” According to Harman, the one who looks to rationality to rein in the profusion of differing moral beliefs only succeeds in doing so insofar as she openly or covertly departs from Science and packages an “unreduced normativity” into her notion of rationality. In other words, appeals to rationality defuse the problem of moral reasons arbitrariness only by abandoning materialism.

If Harman is correct, the resources afforded to moral philosophy within an epistemology dominated by a science beholden to materialism fail to support the kinds of reasons needed if we are to avoid a particularly problematic sort of moral arbitrariness—one that Harman purports to embrace. This kind of arbitrariness, where moral reasons are seen either to be directly the product of non-moral forces unresponsive to moral considerations, or else to be unable to uniquely justify one moral judgment over another (or to suffer both difficulties), is what I am calling moral reasons arbitrariness. And, as Harman himself amply shows, this sort of arbitrariness leaves morality in shambles; witness what he says about Hitler and the guy who murders Ortcutt. To use the imagery from the Introduction, Harman has led us into a coal pit in which we have no reasoned moral criticism of Hitler or Genghis Khan and no reasoned moral approbation for Mother Teresa or Martin Luther King, Jr., all such reasons being explicable in non-moral terms that show

133 Harman, Explaining Value, 91.
134 Though I wonder if an apraxia argument might be lurking in the area.
135 Harman, Explaining Value, 7. There’s also a sad passage on Stalin on the same page.
that the authoritative prescriptivity of moral reasons people take to uniquely justify a course of action owes it *shape to non-moral causes*—leaving nothing that is *uniquely* justified. Whatever one counts as “justified” owes to some group or another whose cooperation one wants for “scientifically understandable” motivating reasons, and the same holds true for what other people hold to be justified. Harman clammers down into the pit of moral reasons arbitrariness. In effect, Harman says there is no other road, so welcome to the pit; perhaps it will feel more homey if we can dress it up a little.

In effect, my argument in this dissertation suggests that Gibbard, Korsgaard, and Post fail to identify roads that would allow us to avoid the same sort of arbitrariness. Upon close examination, I argue that they end up roughly where Harman does, their best efforts to underwrite some sense of justification in “ethics proper” without appealing to authoritatively prescriptive moral reasons (or at least a good stand-in) from within the metaphysics they share with Harman coming to naught. Willy-nilly, they end up chewing gingerly the same bullets Harman chomps down on.

That said, Korsgaard’s worry about the “*normative or justificatory adequacy*” of our moral reasons is well placed and of vital importance; we should not follow Gibbard into the coal pit without utterly compelling reasons for doing so. Though I will argue in Chapter 4 that Korsgaard herself lacks the resources to avoid moral reasons arbitrariness, her instinct about the problem is on target, and her insistence that justifying reasons do not collapse into motivating reasons is essential to avoiding the problem. To find that the realm of external reasons with the authority to rightly prescribe some courses of action over others cannot be

---

maintained in its own terms would be a significant set back for the human business of living well together, greatly impoverishing the resources we have for reasoned moral critique, for resisting the tyrannies of majorities (be they in 1930’s Germany or the Antebellum South), and doing what is right when it gets hard. Indeed, in a very significant way, if Harman is right, we would be left to wonder if there is a right at all.

Let’s start then with Harman’s first objection to the distinction between justifying and motivating reasons, namely, that the distinction is “an ad hoc manœuvre” and that no reasons have been given for distinguishing between justifying and motivating reasons.\footnote{As noted above, Harman still has a concept of justification, but it is tied to “implicit agreements” that people make with groups of other people whose cooperation they need. Without motivations to endorse what others endorse there are no justifying reasons.}

One finds it a bit difficult to believe that Harman means to say this. Let’s start with moral experience.\footnote{Of course I am far from the only philosopher to appeal to the phenomenology of moral reason-giving when replying to one reductionistic materialism or another. “There is,” as Charles Taylor notes, “a tension between phenomenology and [materialist] ontology. The former, properly and honestly carried through, seems to show that values of this higher status...are ineradicable from our deliberations of how to live. But ontology, defined naturalistically, says that properties of this kind can have no place in an account of things in the world” (Taylor, “Ethics and Ontology,” 310).} Surely it’s quite common to find people offering up reasons that they think justify their actions as such. These reasons are not couched in the terms of motivations, but rather in the authoritatively prescriptive language of “ought,” or “should,” or “have to,” or “it’s right to.” Now perhaps the distinction can be explained away somehow or other, but it is in fact a distinction that many—if not all—people make all the time. Moreover, it seems to be a common experience of many that motivating reasons are opposed by justifying reasons. I don’t know how much more common you can get than a country ballad, so let me quote the words of a Randy Travis song from a few years back.
On one hand I count the reasons I could stay with you
And hold you close to me, all night long.
So many lover’s games I’d love to play with you
and on that hand I see no reason why it’s wrong

But on the other hand, there’s a golden band
To remind me of someone who would not understand
On one hand I could stay and be your loving man
But the reason I must go is on the other hand.\textsuperscript{139}

On the one hand, our love-struck country crooner finds powerful motivating reasons,
supplied by sexual and perhaps emotional drives in his SMS, to cheat; but, on the other
hand, he considers the reasons that address him with what he must do or should do or
believes it is right to do, namely not break the solemn promises he made to another person.
Now Harman assumes that what supplies or is the source of these reasons that stand over
against the clear motivating reasons on the other side is not as clear as it is in the case of the
motivating reasons, and this is in large part what drives Harman’s opposition to reasons that
justify actions (though again, as Scanlon argues, it may be that “the notion of a desire is not
nearly as clear as is commonly supposed”).\textsuperscript{140} But even if Harman were right, it would not
change the fact that this “on the one hand/on the other hand” experience—where people
sense a difference in kind between two sides of the opposition—does match the experience
of many (all?) people.\textsuperscript{141} The distinction between justifying and motivating reasons does in


\textsuperscript{140} Scanlon, \textit{What We Owe}, 77.

\textsuperscript{141} It seems to me that this does in fact include all adult human beings who are not mentally impaired, but I don’t want to barnacle my argument with side issues not relevant to the main thread of the argument. That said, the philosophers who deny this seem to me to be “in the grip of a theory,” as they say; and the
fact get drawn with great regularity by many people. Perhaps the distinction can be explained away, but the fact remains that the loss of the distinction, or the coming to believe that it is superficial and controlled by forces that must be described as arbitrary from within “ethics proper,” would be thought by many people (and certainly myself) to strip their daily lives of an important sort of guidance. This common experience and the widespread sense of its importance would be one very powerful reason to take the distinction between justifying and motivating reasons seriously.

But perhaps Harman means that no “scientific” reasons have been given for the distinction, and here we’ll jump down to his third reason for being skeptical of the distinction between justifying and motivating reasons. Indeed, this seems to lie at the root of Harman’s rejection of justifying reasons. Questions of how science materialistically construed and our lived experience fit together tap into a huge issue—Thomas Nagel built a very successful career around it—and I will not pretend to resolve such questions here. I will be delighted if my few brief suggestions give pause for those who think this is a closed question. Indeed, one of Nagel’s great virtues is that he insisted that there are grounds for real puzzlement here, a puzzlement that should not be artificially foreclosed. Nagel rightly insisted that the tensions between science as understood by materialists and lived experience not be glossed over, even if—as I think—he sometimes failed to see just how deep those tensions run and the force they exert for loosening the materialist stranglehold on science.

anthropological examples of various peoples who do not make a “moral” distinction between right and wrong, leave me rather cold, for it does seem that they have a sense of things being “out of order” or “not to be done” in some fairly robust sense—a sense that is a pretty good approximation of the core of what moral language seems to be about. That said, I do not wish to make an issue out of this; I only insist on the “many” that is in the text and which cannot be gainsaid. And that is enough to make the point I need to make.
At least in part, science is a useful, artificial exercise we set ourselves to see how far we can go in explaining things when we limit ourselves to certain terms—materialist terms. When we reach the limits of that experiment—as we undoubtedly do (think of *qualia*, intentional action, normativity, first causes, Heidegger’s *Seinsfrage*, free will, and so on)—we who have seen the usefulness of the experiment are sorely tempted to either dismiss the limits as temporary rather than insuperable, or else simply to eliminate whatever cannot be made to fit the terms of the game, taking up an “eliminativism” about the phenomena in question. But what entitles us to conclude that those materialist terms of the experiment are the terms that limit the object of its inquiry? If we find parts of our experience that defy explanation in those terms, it is not necessary that we conclude that those experiences are illusions or epiphenomena; perhaps we have reached the end of the explanatory game we were playing. Perhaps it indicates the need for an expansion of our explanatory terms. Maybe Harman knows of some good argument why, when the terms of materialist explanation are exhausted, everything for which we cannot account in those terms must be read out of our experience or regarded as in principle explicable in those materialist terms rather than broadening our explanatory terms. But I do not know of any such argument, and I’m pretty sure some powerful objections could be raised against an argument that purported to show this. At root the limits on reality that Harman observes appear to flow out of a commitment to a particular metaphysic rather than from good arguments. Perhaps people bring materialism to science rather deriving materialism from science.

Thomas Nagel puts his finger on the issue when he observes that many in the academy today have what he calls a “cosmic authority problem” in which they actively hope...
that the universe doesn’t include anything like God or the soul or forms or any other
“spooky” things.\textsuperscript{142} It is this cosmic authority problem, Nagel avers, “that is responsible for
much of the scientism and reductionism of our time.”\textsuperscript{143} This cherished hope lies behind
the refusal to acknowledge that science runs into explanatory limits, a refusal expressed by
issuing wild promissory notes for reductions (at least of a token-token variety) not yet
performed or eliminating that which is obviously not of the right kind to be explained in the
requisite manner. Yet the aporias remain. Here’s how Nagel puts the matter: “This is the
general form of all failures of reduction. The perspective from inside the region of
discourse or thought to be reduced shows us something that is not captured by the reducing
discourse.”\textsuperscript{144} Whenever we attempt to complete the reduction, the “I”—complete with all
its phenomenal consciousness, intentionality, normativity, and so on—remains in the
position of the one who is doing the observing, and is never encompassed by the reducing
glare.\textsuperscript{145} It is always on the other side, and with it are all the lived experiences of the color of
a sunset and how it feels to be kissed for the first time, the experience of knowing that one
should not have cheated on an exam, and on it goes. Faced with a systematic failure of
explanatory reduction, the insistence that materialist reasons must yet be given, is a demand
arising from a faith that a certain metaphysic must be true. But it’s precisely the failure of

\textsuperscript{142} Thomas Nagel, \textit{The Last Word} (Oxford: Oxford University Press, 1997), 130. John McDowell uses
the language of “spookiness” as a cute way of voicing his materialist credentials. See John McDowell, \textit{Mind and

\textsuperscript{143} Nagel, \textit{The Last Word}, 131.

\textsuperscript{144} Nagel, \textit{The Last Word}, 73.

\textsuperscript{145} This is not unrelated to the distinction between a “problem” and a “mystery” that is central to
much of Gabriel Marcel’s thought. See Gabriel Marcel, \textit{Being and Having: An Existentialist Diary}, trans.
explanation that shows that the demand is an unreasonable one flowing out of a metaphysical commitment. The failure of explanation might be better met by understanding that the experiment has reached its limits and that broader terms of explanation are called for.

As I said, this issue is much larger than anything I could hope to resolve in a whole book dedicated to the topic, much less in a couple of paragraphs of a dissertation (and I reemphasize that I do not need to resolve the issue here). I do think, however, that these suggestions, combined with the clear direction of our experience, should give one pause before plunging ahead with Harman. As Nagel observes, “any attempt to account for one segment of our world picture in terms of others must leave us with a total world picture that is consistent with our having it… A proposed reduction in any of these domains must be powerful enough to either accommodate or overcome what we think we know from inside them.” The reduction of authoritatively prescriptive moral reasons to motives does not have anything like the explanatory potency to overcome what many people know from the shape of their experience, nor can it accommodate that lived experience. Harman’s demand for answers conformable to his materialist metaphysic certainly is not mandatory. Perhaps it is even epistemically strained in its eliminativism.

So far I have spoken to Harman’s first and third reasons for rejecting the distinction between justifying and motivating reasons. What of his second reason, his worry that the

---


147 Nagel, *The Last Word*, 74.
distinction might lead to a regress? Now it’s not clear why this should be any more problematic for authoritatively prescriptive moral reasons than it is for motivating reasons. Apparently Harman thinks it obvious that for motivating reasons the potential regress comes to an end in a fact about our human nature. Perhaps. But then why couldn’t the potential regress for authoritatively prescriptive moral reasons resolve itself in a fact of reason or some other inherently justifying reality? Undoubtedly, Harman would say that this is spooky, and “queer,” and inexplicable.\textsuperscript{148} Two things should be noted here. First, it’s not clear that we really understand the motivational end point any better. Certainly we’re more accustomed to talking about motivation as tied to human nature, but it’s not clear what exactly this talk specifies. Second, at this point the objection really collapses into the third objection, which is what really drives Harman’s project anyway.

I conclude, then, that Harman has not offered compelling reasons for being skeptical of the distinction between authoritatively prescriptive moral reasons and motivating reasons—certainly none that merit overturning the weight of such a central element of the experience of many (all?) people. Mostly he depends (as he no doubt can) on the sympathy most of his readers will have for his metaphysical commitments. But these are pretty thin grounds for such a momentous dismissal—a dismissal that goes in the face of the lived experience of many people. Harman’s reasons for overturning the natural and intuitive distinction between motivating reasons and authoritatively prescriptive moral reasons that people understand as uniquely justifying a particular course of action are very thin, primarily consisting of his commitment to a particular metaphysical stance. Some

\textsuperscript{148} This kind of objection is closely allied with J.L. Mackie’s “argument from queerness.” See Mackie, \textit{Ethics: Inventing Right and Wrong}, 38ff.
materialists will want to dispute his understanding of what their shared metaphysical commitment implies for moral reasons, and hold on to both the metaphysical commitment and the authority of moral reasons that justify actions. Indeed, in their different ways, Gibbard, Korsgaard, and Post each try to do this. For my part, I think Harman is right about what a commitment to materialism does to authoritatively prescriptive moral reasons, and that we have a good reason (one of many) to question the metaphysical commitment (though it is not the purpose of this dissertation to argue for that point). In any case, the many moral philosophers who have concerned themselves with issues of moral justification are right to have done so; they are not shadowboxing.

This much is clear, I believe: from the standpoint of “ethics proper,” Harman’s philosophy is highly problematic, eliminating independent, authoritative prescriptivity from moral reasons and leaving us with the realization “that, for Hitler, there might have been no reason at all not to order the extermination of the Jews.”\(^{149}\) In short, in Harman we have an example of one philosopher who starts with materialism and ends in the coal pit of moral reasons arbitrariness (the problem I am fleshing out in the present chapter). I suspect it is not easy (if possible at all) to avoid such moral reasons arbitrariness given the metaphysical hobbles Harman is wearing, and subsequent chapters of the dissertation will explore whether or not Gibbard, Korsgaard, or Post (Millikan) have devised a way to avoid this coal pit while accepting the same metaphysical assumptions.

C. Some Test Cases

Before leaving Harman behind, I want to draw out two challenges that he poses that will serve as test cases in the following chapters for the way Gibbard, Korsgaard, and Post attempt to give an account of moral reasons. These will be (1) the case of the contented criminal, and (2) the case of the viable dictator, both of whom we met in a passage quoted above. I will amplify Harman’s thoughts on these issues by drawing on G.A. Cohen’s critique of Christine Korsgaard, and Richard B. Brandt’s important work in moral philosophy.150

First, let’s look at Harman’s “contented employee of Murder, Incorporated” who finds no good reason not to murder someone. Harman takes it that for someone who “was raised as a child to honour and respect members of the ‘family’ but to have nothing but contempt for the rest of society,” it would be pointless for us to appeal to reasons to convince him not to murder somebody.151 He really doesn’t have any such reasons. But this isn’t just a hypothetical case, Harman argues. There are criminals who in fact have no compunctions when it comes to murder. Furthermore, “their lack of acceptance of the relevant principle does not appear to rest on a failure to notice certain facts or incorrect reasoning or any failure to follow out certain reasoning. Nor is it always because of stupidity, irrationality, confusion, or mental illness.”152 In short, Harman believes that there are people who are fully rational and who are fully aware of the relevant facts, and who yet


151 Harman, Explaining Value, 5.

152 Harman, Explaining Value, 33. See also 40, 86-7.
have no reason not to murder someone else. There are only motivating, internal reasons in the realm of morality, and some people find no reason internal to themselves that motivates them not to murder. Given their implicit agreement with “the Family,” the reasons they have point toward murder in some situations. Nor are there rational grounds for criticizing them—we can only hope that there aren’t many such people, and that there are enough of “us” who do find we have compunctions against out-group murder that we will be able to make it difficult for such people to murder others. But, the point remains: “it can be ‘rational’ for different people to accept different moral demands ‘all the way down’.”\textsuperscript{153} If Harman is right, when moral reasons are restricted to motivating reasons, a fully rational, fully informed person can have no reason whatsoever not to murder another person.

This poses a direct challenge to philosophers who wish to derive moral reasons from the rationality of contingent rational beings, and later chapters shall consider just how much of a problem the contented criminal poses for Gibbard, Korsgaard, and Post. This challenge is particularly sharp for those who, like Harman, limit our moral reasons to internal reasons. Thus, we find G.A. Cohen pressing a related type of case—the case of the “idealized Mafioso”—against Christine Korsgaard’s attempt to formulate justifying reasons from the resources internal to a person’s “practical identities.” We will see that these sorts of cases can be used to great effect against the theories of practical reasoning and morality developed by Korsgaard, Gibbard, and Post—and, by extension, those theories that are relevantly similar.

\textsuperscript{153} Harman, \textit{Explaining Value}, 33-4.
In the absence of reasons that can arbitrate our discussion with a contented criminal (or similar type), the only thing “we” can do is try to make it practically difficult for “them” to act on their reasons, trying our best to inculcate others in our society with our preferred way of doing things and hoping that they will not be swayed by other forces. As Harman puts it, “we might not be able to provide the assassin with the slightest reason to desist unless we were to point to practical difficulties, such as the likelihood of his getting caught.”154 In short, “we” create as many practical difficulties as we can for those who disagree with “us.” This sort of practical difficulty a person encounters in pursuing her self-interested ends is a matter of what Richard Brandt calls “viability.” Roughly, a moral code is “viable” for a person if she can get others to go along with it, one way or another.155 “Thus all selfish rational persons, except for magicians and dictators, will support a minimal moral code for selfish reasons, and the code they choose will maximize the expectable welfare of the group.”156 But what about this viable dictator? Do we have anything to say to him? Are there any reasons for someone who is reasonably sure that he has correctly calculated that he can viably install himself as a dictator not to follow this path? But we can make the problem even more pointed (and realistic) by remembering the lessons that Harman would have us learn. Such a dictator in making also need not be concerned with groups of people who are irrelevant to achieving the ends that interest him. Such people may be left out of account in calculating viability. Moreover, one’s calculations of viability may also include considerations of how much others may be cowed, misled, bamboozled, put off until they

154 Harman, Explaining Value, 5.


can be managed, etc. One begins to realize that the “viable dictator” is not a simple “yes or no” proposition; there can be degrees ranging from Hitler (who was viable indeed—had he played his cards differently with Russia and England and developed “the Bomb” first—it’s possible that he would have ruled the world until very recently) and Mao, to the petty corporate executive or the tyrant husband who preys on his wife’s insecurities and vulnerabilities, having correctly calculated that his actions are viable. The question, of course, is “Do theories that construct morality from the rationality of contingent rational beings have sufficient resources to constrain the very real person of the viable dictator?”

These are not ethereal or contrived problems (think barn-façades) devised in the philosopher’s armchair to trip up another philosopher’s theory, but rather the real stuff of people’s moral lives. There are many things that are viable for us that we must decide whether or not we should do them. I would guess that I could viably cheat on my taxes, and I would really like to drive the new little Mazda 3 with a sunroof that I might be able to buy with the money thus obtained. But is it right to do so? How is this rightness or wrongness to be understood? What is its source? Moreover, there are real questions about whether or how we might have non-arbitrary grounds for critiquing the person who is quite content with choices that “we” deem immoral. Such questions—questions of whether we should do things that we viably can, and questions of the grounds for critiquing the choices of people who find no reasons not to circumcise women or cheat on their taxes, who are content with the reasons they find motivating them—are questions of our daily lives. A successful moral philosophy ought to address them convincingly.
The following chapters will put pressure on the materialist moral philosophies of Gibbard, Korsgaard, and Post in part by pressing the cases of the “viable dictator” and the “contented criminal” and seeing how they can handle these cases. Failure to handle these cases would go some way to showing how these moral philosophies succumb to the problem of moral reasons arbitrariness. Unlike Harman—who simply embraces the viable dictator and the contented criminal (not merely as theoretical possibilities, but as people like Hitler and many walking our streets), saying that they really have no reason not to pursue what they are motivated to pursue—most materialists think better of this and seek to find a way to avoid moral reasons arbitrariness. Where Harman states that “the issue between relativism and absolutism comes down to the dispute between naturalism and autonomous ethics,” Gibbard, Korsgaard, and Post attempt to formulate a third option. As Harman sees matters, materialism is naturally linked to moral relativism. And any third option needs to stand up under pressure from real-world cases like the dictator who really is viable and the criminal who really thinks he is content. If one materialist moral philosophy after another folds under such pressures, one might begin to wonder if Harman is right to see a linkage between materialism and moral reasons arbitrariness.

In the pages of this dissertation, I have foreclosed any general argument that would attempt to establish a linkage between materialism and moral reasons arbitrariness. My goal is more modest: to make plain the problem of moral reasons arbitrariness and then to argue that three important materialist moral philosophies do in fact succumb to this problem. Harman, however, gives food for thought. Perhaps investigations like those pursued here

---

open up into something larger. Be that as it may (or may not) be, Harman has helped us see a serious problem in moral philosophy: moral reasons arbitrariness. The last section of this chapter attempts to sharpen that problem.

III. The Problem of Moral Reasons Arbitrariness

The problem of moral reasons arbitrariness is widespread in moral philosophy today, it will be suggested below. Indeed, this dissertation is aimed at showing that this problem infects many more moral philosophies than just the few that embrace it outright, as Harman does. But the problem in view here must be distinguished from another kind of “arbitrariness” in moral thought that is not problematic, but which may even be a condition of moral agency. After briefly drawing this distinction the problem of moral reasons arbitrariness will be drawn to a sharp point.

A. Two Kinds of “Arbitrariness”

It is not clear how a person’s will could be determined by authoritatively prescriptive moral reasons with regard to what that person will accept or endorse. One could be exposed to a large number of such moral reasons to φ, reasons that do in fact justify φ-ing, and yet refuse to accept or endorse φ-ing. The authoritatively prescriptive reasons can in fact exist and be presented to an agent, and yet that agent can refuse to accept those reasons. While moral reasons influence the will, exercising their sway through the agent’s recognition of what is justified by those reasons in a particular situation, these reasons do not determine the will. It is always open to the agent to choose otherwise. This sort of “open-endedness” of
an agent’s decision to act upon authoritatively prescriptive moral reasons I shall call 
\textit{deliberative open-endedness}. It is, perhaps, a kind of “arbitrariness” in the moral lives of human 
beings, but it is important to recognize in it the condition of human free will. As such, it 
should be welcomed, even if it is a source of much of the human moral predicament about 
what we should do and why we should do it. A definitive answer to these questions that 
adopts of no possible rejection by human beings simply does not exist. If I am right, this is 
enemic to our moral life as human beings.

The other kind of arbitrariness is much more problematic, having to do with 
whether all authoritative moral prescriptivity finds itself undermined by an arbitrariness that 
owes to the insufficiency of the ontological provenance of justifying reasons. I shall call this 
kind of arbitrariness, \textit{moral reasons arbitrariness}. I will consider moral reasons arbitrariness 
shortly; first let’s turn our attention to the question of deliberative open-endedness.

\textbf{B. Deliberative Open-Endedness as an Unavoidable Feature of Human Morality}

John Hare’s book, \textit{God’s Call}, makes a convincing case for a fundamental role for 
endorsement or acceptance in our moral life. It can be tempting to overreact to the notion 
of human endorsement or acceptance as constitutive of morality that one finds among 
noncognitivists, but John Hare and his father, R.M. Hare, argue convincingly that some 
place must be found for endorsement as we think about morality. If we fail to do so, we 
leave out the crucial moment of decision about what to do. And yet, as John Hare rightly 
stresses, this moment of decision need not—indeed, should not—be thought of as constituting 
moral rightness and wrongness. Instead, it is possible that human moral decisions can be a
matter of recognizing or acknowledging or submitting to authoritatively prescriptive moral reasons that have their authoritative standing independently of their being (or failing to be) endorsed by us. They can, in short, be objectively or authoritatively prescriptive even when no human being endorses them. The acceptance or endorsement critical in moral judgment does not bear on the existence of authoritatively prescriptive moral reasons that determine whether or not we have endorsed that which we ought to have endorsed.

That the moment of decision in an agent’s moral action is a moment of endorsement or acceptance is strongly emphasized by R.M. Hare in *The Language of Morals*. When an agent is pondering what should be done in a particular situation,

It is left to the decision of the agent whether to act upon the principle or not. Now to use the word “inference” of a procedure like this is seriously misleading. When someone says, either “This is false, so I won’t say it”, or “This is false, but I’ll say it all the same, and make an exception to my principle”, he is doing a lot more than inferring. A process of inferring alone would not tell him which of these two things he was to say in any single case falling under the principle. He has to decide which of them to say.\(^1^{158}\)

Or again, in talking about how one could answer someone who kept pushing one “But why *should* I live like that?” question after another, seeking fundamental bedrock for why something should be done, R.M. Hare answers that after having given a “complete specification of the way of life” from which a should statement arises,

We can only ask him to make up his own mind which way he ought to live; for in the end everything rests upon such a decision of principle. He has to decide whether to accept that way of life or not; if he accepts it, then we can proceed to justify the decisions that are based upon it; if he does not accept it, then let him accept some other, and try to live by it.\(^1^{159}\)


\(^1^{159}\) Hare, *The Language of Morals*, 69. Wittgenstein’s influence is clearly on display here.
Decision, and the commendatory force flowing from it, “is of the very essence of morals,” according to Hare.\textsuperscript{160} As Hare sees things, it is an individual’s acceptance of a particular application of a moral principle and, ultimately, of the complete form of life that makes sense of it, that constitutes morality. The individual’s decisions are crucial.

One of Hare’s central concerns in\textit{ The Language of Morals} is to expose the flaws in any “naturalistic” theory that would ignore the critical issue of where the imperative element in morality comes from, deriving morality from facts and inferences from those facts. Hare locates the critical imperative element of morality in persons and our acts of endorsement. Indeed, it is hard to see how this imperative element, which Hare rightly sees as crucial, could be derived from anything that was not a person. In giving this account in\textit{ The Language of Morals}, however, Hare lacks a satisfactory account of how human endorsement and prescription is constrained, and thus he succumbs (I would argue) to the problem of moral reasons arbitrariness.

That said, Hare is right to emphasize the importance of human decision in morality. Finally, an agent does have to decide what she will accept, and that decision is not determined by the authoritatively prescriptive moral reasons relevant to the situation she is in. Even after all those reasons are available to the agent, what she will endorse is still up to her.\textsuperscript{161} Becoming “morally adult,” R.M. Hare states, happens when we “learn to use ‘ought’-sentences in the realization that they can only be verified by reference to a standard or set of

\textsuperscript{160} Hare, \textit{The Language of Morals}, 54.

principles which we have by our own decision accepted and made our own.”\footnote{Hare, The Language of Morals, 77, 78.} This is true; but if there be external moral reasons, it is nonetheless the case that the principles thus accepted may fly in the face of the reasons that are there whether we recognize them or not and thus may be mistaken. Sincere moral judgment does not automatically render the moral content of the judgment justified.

The gap between sincere moral judgment and moral justification is made much clearer in John Hare’s work, particularly in his book, God’s Call. In this book, John Hare attempts to reconcile elements of what he calls “moral realism” and “moral expressivism,” drawing them together into his own position, which he calls “prescriptive realism.”\footnote{Hare, God’s Call, 1. See also the chart on page 2 of the same work.} Apart from a reading of Kant’s moral philosophy at the end of the book that leaves me completely cold,\footnote{John Hare is a central figure in a movement among some philosophers of religion and theologians to reclaim Immanuel Kant as an important inspiration for thinking about Christian philosophy and theology. This movement has received recent expression in a collection of essays, in which an essay by John Hare features prominently: Chris L. Firestone and Stephen R. Palmquist, eds., Kant and the New Philosophy of Religion (Bloomington, IN: Indiana University Press, 2006). Hare’s reading of Kant, which is central to his important book, The Moral Gap: Kantian Ethics, Human Limits, and God’s Assistance (Oxford: Clarendon Press, 1996), and is also elaborated in chapter three of God’s Call, and some recent essays (John E. Hare, “Kant on Recognizing Our Duties as God’s Commands,” Faith and Philosophy 17, no. 4 (October 2000): 459-78; and John E. Hare, “Essay Review: Christine Korsgaard, Creating the Kingdom of Ends,” Faith and Philosophy 17, no. 3 (July 2000): 371-83), strikes me as wildly optimistic about the usefulness of Kantian themes for Christian thought, and just plain mistaken on a number of points of Kant exegesis. For a salutary reminder of why Kantian philosophy is a reed that will splinter in the hand of the Christian philosopher or theologian who leans on it, see Keith E. Yandell, “Who Is the True Kant?” Philosophia Christi 9, no. 1 (2007): 81-97.} I find the general structure of Hare’s arguments to be well thought-out and correct in important ways. John Hare corrects his father’s overemphasis on the role of decision in human morality, while retaining the important insight that “moral judgment is prescriptive,
expressing the will.”

But in John Hare’s work, this insight is balanced with the recognition that when we commend “we are not putting something into the world that was not there before.” Rather, there is something that commends actions to us, “a call that deserves our obedience.” Again: “To judge that a thing is good is not merely to report the magnetic force, but to judge that the thing deserves to have that effect on us.” If this is right, then commendation or prescription can be in response to recognition of a rightful, authoritative prescriptivity that comes to expression in reasons that address us whether we wish to be aware of them or not. The decision is not taken from us; we can acknowledge or refuse that which is there anyway. But John Hare insists that it is possible for our moral judgments to be responses to authoritatively prescriptive moral reasons that exist quite independently of whether or not we acknowledge them in our judgments. This corrective to the excessive importance sometimes attached to human endorsement or acceptance is of first importance.

---

165 Hare, God’s Call, 8. It seems to me that sometimes John Hare’s exegesis of his father’s work softens R.M. Hare’s “antirealist” tendencies more than is merited, at least with respect to the elder Hare’s influential early work. The R.M. Hare of The Language of Morals is indeed a “creative antirealist,” despite John Hare’s protests (God’s Call, 15). This is particularly clear with respect to the situation of those whom we might call (with a nod to Kripke and Putnam) inaugural prescribers, that is, those who first set the moral kinds inhering in a form of life. In accounting for the descriptive content of moral judgments as arising from the “general acceptance” of the principles on which they rest (The Language of Morals, 195), R.M. Hare insists that these principles “cannot have been informative from the beginning. The standard must have got established by some pioneer evaluators… The standard became established by… making commendatory judgements which were not statements of fact or informative in the least” (The Language of Morals, 146-7; see also 118-9, 136, 150). It is clear that the arbitrariness problem faced by the “inaugural prescriber” or “pioneer evaluator” is also faced by the “reforming prescriber.” Indeed, a strong case could also be made that the ordinary prescriber within the view laid out in The Language of Morals would also face this quandary in a fairly strong form. It is interesting to compare Hare’s views here with the views of Max Weber on the “routinization of charisma.”

166 Hare, God’s Call, 39; see note 25, above.

167 Hare, God’s Call, 38.

168 Hare, God’s Call, 21-2; emphasis in original.
At the same time, it is important to recognize, as both R.M. Hare and John Hare do, that our moral endorsement or acceptance of a course of action can always run counter to authoritatively prescriptive moral reasons. This deliberative open-endedness leaves us free to fly in the face of the weight of the moral reasons, ignoring them for reasons of our own. The human being is a free moral agent who cannot be stripped of the responsibility for his own actions.

Before moving on to lay out the problem of moral reasons arbitrariness, I would like briefly to say how I believe deliberative open-endedness fits together with the judgment internalism I endorsed above. Judgment internalism, one will recall, is the position that once a moral judgment is made and a course of action is endorsed, there is a very strong motivation to follow through—though even this can be overturned. This may initially seem to be in tension with the deliberative open-endedness I am also arguing for. But brief reflection on the matter shows why this is not a difficulty. In short, the issue of the decision that is the concern of deliberative open-endedness is prior to the issue of endorsement in a sincere moral judgment that is the concern of judgment internalism. It is only once the decision has been made to endorse something through a sincere moral judgment that the motivating force attached to that judgment kicks in and urges us to go ahead and act in the direction thereby endorsed.
C. Moral Reasons Arbitrariness

Let’s return to Korsgaard’s worry. She illustrates her worry by sketching “a moral theory that gives morality a genetic basis,” a view she calls “the evolutionary theory.”\[169\] It’s a standard sort of view of the type on offer by various “sociobiologists,” where human moral behavior is explained by reference to drives toward the preservation of the species that were selected due to the fact that they helped our ancestors reproduce at a higher rate than related organisms that did not behave in this manner. Perhaps this adequately explains the behavior of the species. But Korsgaard presses us to ask “whether if you believed this theory, it would be adequate from your own point of view.”\[170\] In other words, where would you stand with a particular moral obligation if you had come to believe that it was fully explicable in the terms on offer by the sociobiologists: “Morality has no basis other than the way that we feel. Morality is an illusion of the genes put in place to make us good social animals.”\[171\] If you believed that, Korsgaard wonders, then how would morality be justified to you as a person—especially when the requirements of morality were costly to you personally (as they often are)? What would become of the authority of its prescriptions? Having in hand an evolutionary explanation of the pull of morality on you and the reason why it feels justified, you might sensibly take yourself to have good reason to resist the pull of the moral “should” or “ought.” From the point of view of moral justification, the explanatory terms may well be taken to undercut the authority that is thought to attend the idea of

\[169\] Korsgaard, Sources, 14.

\[170\] Korsgaard, Sources, 14; emphasis in original.

something’s being “right” or “obligatory” or “justified,” these ideas now being accounted for in explanatory terms that can only seem arbitrary from the point of view of moral justification. If I come to believe what I have been calling my “moral obligation” is nothing more than a feeling that wells up in me due to the accidents of my evolutionary history (or my cultural background, or what I ate for dinner, or whatever), then—from the point of view of moral deliberation itself—those feelings are arbitrary and lose their authority over my deliberations on what I should do in this particular situation. In a world explained in such a way that the only address to me as a moral agent comes from myself and (perhaps) from other contingent personal beings, and where even the pull of this address is finally to be explained in other terms, I am left to cast around for other terms in which to couch my moral deliberations—or else to find some way of rehabilitating the terms that have come to seem arbitrary.

That is what worries Korsgaard in her example of the “evolutionary theory” of morality. Korsgaard has done us a real service in thematizing the “Why be moral?” question as a justificatory question (as John Hare points out on different occasions), and I think that her “evolutionary theory” example and her reflections on the problem it raises for us are fruitful for moral philosophy. In particular, Korsgaard is right to focus on what she calls the problem of “justificatory adequacy.” An explanation is not enough for us as moral agents.

172 Loyal Rue draws exactly this conclusion from the work of the sociobiologists, and he proposes that the few who are able to look this reality full in the face and not succumb to moral nihilism and despair should work to formulate a “noble lie” for the masses so as to prevent the dissolution of society, which can operate smoothly only when most people are operating under the delusion crafted by evolution for our own good. See Loyal Rue, By the Grace of Guile: The Role of Deception in Natural History and Human Affairs (Oxford: Oxford University Press, 1994).

“When we do moral philosophy,” Korsgaard observes, “we also want to know whether we are justified in according [practical] importance to morality… We want to know what, if anything, we really ought to do.” Moral agents need justification in deciding what should be done, but when a person believes an explanation that would account for morality in non-moral terms, this seems to render morality optional rather than obligatory from the person’s own point of view. In making moral decisions, people consider and weigh support relations. A moral philosophy that fails to address people in this deliberative capacity or that undercuts those deliberations by holding that deliberations about support relations are not what they seem to be, but are a vehicle through which non-moral forces operate on a person and the people around her—such a philosophy falls prey to moral reasons arbitrariness.

Let’s introduce a toy example to begin focusing the problem. Why should you avoid telling a lie if you know that the explanatory reason for why you feel that it is not justified for you to lie is that the burritos you had for lunch didn’t agree with you? If you actually believed that your burritos completely explained the feeling and your tendency to be swayed by any particular chain of reasoning in support of it, and that—had you eaten a light salad instead, you would have felt no such compunction against lying and that the Kantian casuistry would not have had no effect—the ability of those things to justify to you a course of action as you deliberate about what to do would be undermined. There is no reason not to say, “Screw it. I should have had the salad so I wouldn’t have felt this way. Those

---

174 Korsgaard, Sources, 13; emphasis in original.
burritos are always so disagreeable… I’ll go ahead and lie.” Believing the explanation brings it into one’s justificatory ruminations.

Now, clearly the example is silly, but it serves to make a point. If you believe the explanation, you at least think it leads your moral reasons around by the nose; according to what you believe, they only seem authoritatively prescriptive because of what you ate for lunch. Things just happen to be the way they are, but they would have been quite different, you believe, had certain non-moral factors determined your feelings of the justifiability of your performing such-and-such an action in light of (what you had taken to be) independent, authoritatively prescriptive moral reasons. Your belief in the explanation forces the non-moral explanation into the heart of your justificatory deliberations about what should be done, where it shows up in such a way as to undercut your belief that something has the independent, authoritative prescriptivity characteristic of genuine moral reasons. You yourself, at any rate, believe that sense of there being an independent, authoritatively prescriptive moral reason not to lie would have been different had you eaten the salad. For you as a person, considering the support relations involved in a moral situation, the explanation you believe removes precisely the support relation, casting it in different, non-moral terms. Your deliberations are a sham, operating at the behest of non-moral forces rather than support relations that should be taken into account. So why not just discount the (now explained) sense of not lying being justified by those “reasons,” especially where it would be costly to do otherwise and tell the truth? Given the explanatory story, and if there is not some special additional story to be told, there is (as you believe) nothing that uniquely justifies the one course of action over the other in light of authoritatively prescriptive moral reasons. The explanation you accept
sweeps away the support relation as such, exposing its justificatory power as really one more vehicle for the operation of non-moral forces. The problem is that you take it that—had the non-moral forces been different—some other belief would have shown up in your moral deliberations seeming to be uniquely justified by the authoritatively prescriptive moral reasons.

Let me give one more contrived example, before moving on to a real world case in the person of Gilbert Harman. The fact that I felt dyspeptic after downing too many chocolate malts at the local ice cream parlor might explain what motivated me to yell at my children, but it fails to address the question of whether or not I was right in doing so. Perhaps I am normally lax in my parenting and I don’t correct their cruelty to cats, but this time—because of my dyspepsia—I put an end to it. What I do might be morally justifiable, but it would be hard to argue from my dyspepsia to that justification. It’s just not the right sort of reason to provide moral justification; indeed, it’s much like Jim’s leg jerking upward when tapped with a mallet. If asked, someone might reconstruct some sort of explanatory story for the action involving the dyspepsia that would talk about the reasons why the action occurred (much as one might with Jim’s leg jerking upward). But there are two things to note: (1) the reasons-talk would be on the side of the explanation of the events rather than in the events themselves, and (2) if I believed that my discipline of my children resulted from my dyspepsia, I would “see through” the reasons I might give for my actions to the causes that I would take to be really operative in the situation. In my own eyes, the causal story I believe undermines the justificatory force of the reasons I took myself to have. Nothing is uniquely justified by what I took to be my reasons in the situation, my “reasons”
being otherwise explained in terms of an efficient causality that strips the reasons of any independent, authoritative prescriptivity in my eyes. If all the reasons we have are of a relevantly similar sort, then we are left in a situation of moral reasons arbitrariness. It is an interesting fact that someone might still say of my preventing my children from torturing a cat that there were good reasons for me to do so—even if I did not act for those reasons.

The point here can also be put in terms of the sharp distinction between causes and reasons. In putting the point this way, we can draw on a little help from Thomas Nagel: “our point of view—what we accept on the basis of reason—is a set of beliefs about how things really are.”176 “There is,” Nagel observes, “some way the world is,” and we “cannot escape from or rise above this idea.”177 In other words, we will each believe something about how the world in fact is, even if only as seen in our actions. The world is what it is, and each of us believes something about what that is like. “This,” Nagel notes, “is an idea to which there is no intelligible alternative.”178 The point, then, is this: In accepting some such

---

175 I prefer not to speak of a distinction between causal reasons and justifying reasons, since I don’t think causes are, strictly speaking, reasons, but merely causes. As with the case of Jim’s leg, in our attempts to reconstruct a causal story when asked for an explanation, we can talk about causes as reasons (“The reason Jim’s leg jerked upward is that the doctor tapped his knee with a mallet”). But the support relation is not on the side of what went on with Jim’s knee, but only on the side of our attempts to explain what happened in a convincing fashion. To see this, imagine another explanation: “The reason Jim’s leg jerked upward was because he was having a seizure.” We say that this explanation competes with the other, and we can ask which explanation is better. Now we adduce reasons for one explanation over the other, but no one thinks that the reasons are on the side of what was going on with Jim’s knee (unless, of course, someone advances an explanation like the following: “Jim was messing around with the doctor, and jerked his leg upward.”). In our efforts to understand the causal story, we attempt to see which story fits the causes better and advance reasons for one explanatory reconstruction over another.

176 Nagel, *The Last Word*, 88; emphasis in original. The passage continues as follows: “…together with copious acknowledgement that there is a lot we don’t know and perhaps a lot we can never know about how they really are,” (88-9).

177 Nagel, *The Last Word*, 81. Nagel is arguing against “subjectivist” attempts to undercut science.

story as explaining how things are, the elements of that *believed* explanatory story are brought over into the realm in which moral justification happens. In short, they become considerations to be weighed. How the explanatory story you accept bears on your moral reasons becomes a relevant consideration as you deliberate about your reasons to do one thing rather than another. If you believe that the burritos explain why you think there are reasons not to lie, you think that the support relation is undermined. In this way, causes insinuate themselves into the realm where moral deliberation happens and threaten to wreak havoc with moral reasons, exposing them as “really” being vehicles for all manner of non-moral forces or as otherwise not uniquely justifying one moral judgment over another. Thus, if you believe your reasons for not lying really owe to the burritos, the “reasons” lose their ability to support or uniquely justify not lying. Now, perhaps there may be reasons nonetheless, but you will have to go looking for them in a way that either exposes the burrito explanation as bad and not to be accepted or else finds another ground for moral reasons not to lie that can withstand the burrito story that you still accept.¹⁷⁹

Of course the concern here is not with the burrito explanatory story of our toy example, but with explanatory stories that are widely accepted, like the materialist “evolutionary theory” that worries Korsgaard, even as she herself accepts a version of it. I am arguing that the threat to moral reasons is a real one that presents grave difficulties to moral philosophy. We have already seen that a materialist like Gilbert Harman, having

---

¹⁷⁹ Shifting the terms to the real concern of the dissertation, I believe the materialist explanatory story is bad and that another explanatory story can support reasons. Harman accepts the explanatory story and lets go of the moral reasons that have independent, authoritative prescriptivity. Gibbard, Korsgaard, and Post accept the explanatory story, but think there is some additional story that grounds the reasons (or something close enough). It is this last move that is being argued against here, at least in the form it has taken in Gibbard, Korsgaard, and Post.
limited himself to “inner judgments” with respect to morality, sees no way to underwrite the independent, authoritative prescriptivity of moral obligations. Obligations become relative to the subjective motivational sets of individuals, with these arising from the happenstance of their culture, their evolutionary history, and so on. What people take as justified will depend on the groups of people with whom they happen to enter into an “implicit agreement.” Moral “obligations” address a person only as the person accepts them, and both the shape of the obligations that present themselves to a person and the factors that lead to a person’s actual acceptance depend in turn on a variety of non-moral factors. Thus a fully rational person in possession of all the relevant facts might not have any reason not to murder someone for the benefit of “the family.” In Harman’s view, the motivating reasons swallow up the independent, authoritatively prescriptive moral reasons, and he simply embraces the resulting moral reasons arbitrariness. But this, you will recall, is not a tasty bullet to bite: “for Hitler, there might have been no reason at all not to order the extermination of the Jews.”\textsuperscript{180} And the reasons you have not to do so owe to the accidents of your birth that led you into an “implicit agreement” with a group of people who did not think this way. If someone came to believe—really believe—with Harman, that that’s all there is to the story, then it would seem that the weightiness of those reasons would be significantly undermined within that person’s deliberative stance.\textsuperscript{181} Unless there is something more to it,

\textsuperscript{180} See note 89, above. As noted above, this does not mean that any person can just adopt any moral position whatsoever. There will be sociological limits based on the groups of people he finds himself making an “implicit agreement” with. But one may be in Germany in 1937 or in Alabama in 1842. There is a contractarian element to Gibbard’s view, but there is no moral limit on which groups of people one will make an implicit agreement with. This just plays out in accordance with the non-moral forces to which we are subject.

\textsuperscript{181} Now, perhaps this would not sway someone at all with respect to killing Jews (as one would hope, though I remember a news story about some high school boy jeering and laughing when \textit{Schindler’s List} was
an additional story to be told that somehow reinstates the reasons as such, then it’s difficult to see how any belief about moral obligation could be justified from within that person’s deliberative stance as shaped by the explanation she accepts. As she understands it, the independent, authoritatively prescriptive moral reasons that purportedly justify the action owe to the non-moral forces that stand behind their acceptance. If there is nothing more to be said, then the fact that she believes that a practically limitless variety of actions are understood to be justified by different people due to the way they are in the grip of certain non-moral factors, and that all people—including herself—would have thought different courses of action morally justified had the die been cast differently with respect to these non-moral factors, all of this undermines independent, authoritatively prescriptive moral reasons, leaving the agent to deliberate without the guidance of such reasons. All the (incompatible) actions people take to be justified emerge from some causal story or another that is unresponsive to the “reasons” agents take themselves to have, and the actions I take to be justified simply because of the way these forces happen to have gripped me I would have seen differently had different causal forces gripped me. I (believe that I) see that the reasons I take to uniquely justify an action actually provide no such ground; the support relation is cut, seen (or believed) to be a merely causal matter operative apart from any responsiveness to reasons. The incompatible judgments of other people stand on the same justificatory ground I claim for my own, casting grave doubt on the unique justification I had taken to support my moral judgments. What I had taken for a response turns out to have been caused, and what I had taken to be uniquely justified enjoys no rightful claim over

shown at their school), but what about cheating on taxes or bilking investors or cheating on one’s husband with the good-looking, engaging guy at the office, or running off with the younger woman and leaving the old lady with the kids?
other, incompatible judgments and actions—all our “reasons” owing to the operation of the same causal forces that are unresponsive to these “reasons.” In this case, independent, authoritative prescriptivity does not come to expression through the reasons themselves; no action is uniquely justified. Independent, authoritatively prescriptive moral reasons are seen through from within the deliberations of the moral agent herself, leaving her to see that no one course of action is uniquely justified by the reasons that can be recognized in the situation.

The kind of arbitrariness I am highlighting is not unlike one identified by Seiriol Morgan in his critique of the “Cornell realists,” or, as he calls them, the “synthetic realists.”182 The real difficulty for the synthetic realists, as Morgan sees it, is not in establishing a sort of moral realism, but that any “realism” devised in the manner they propose cannot show its unique justificatory credentials that entitle it to be preferred over other possible “realisms” established in the same manner. Basically, Morgan believes we can grant (if only for the sake of argument) that there are certain natural moral facts that establish their realist credentials by carrying their explanatory weight in our explanations of certain kinds of events. For example, we cannot understand why a particular revolution happened without understanding that there was an “injustice” at the root of the uprising. Moral facts (much like mental facts in “non-reductive materialism” in philosophy of mind) simply cannot be eliminated from our accounts of the facts. Thus we have a moral realism. Well and good, Morgan says. The real problem, however, is that a whole slew of competing realisms can ride the coattails of this strategy and we are left with the same basic problem, just in a different register. This is because the heavy lifting is done by human nature and its

182 Morgan, “Naturalism and Normativity.” Morgan uses the appellation “synthetic realists” because he has in mind a family of moral realism slightly broader that the philosophers usually understood as “Cornell realists,” wanting to include, for example, Peter Railton among the “synthetic realists.”
propensity to find certain explanatory patterns indispensible; this is what makes for a
“realism.” We are “moral realists” if we find ourselves unable to dispense with explanations
in terms of “equality” or what is “fair” or of the “injustice” of a situation. “Moral realism”
has no deeper grounds than this explanatory propensity of (at least some) humans, and
nothing stands over against this variable and weak human nature to constrain or direct the
prescriptive choices formed from out of this source. As Morgan notes, the specifically
prescriptive element of the realism arises when people constituted in a specific way find
“moral” explanations necessary when in the presence of certain facts. The moral facts carry
weight in the explanations of people who are constituted so as to find such explanations
necessary when confronted by certain facts. But there is nothing that guarantees that
human nature must be constituted in the requisite manner—or even that it usually is.
Somebody could, Morgan observes, be constituted in a more “Nietzschean” manner so that
this person finds “noble facts” carrying explanatory weight rather than “moral facts.” The
same revolution that one person could not explain without referring to “injustice,” a
different person might not be able to explain without referring to the “decadence” and the
“insufficient ruthlessness” of the aristocrats. This “noble realism,” as Morgan calls it, would
carry exactly the same realist credentials as “moral realism,” and—what is important—its
justificatory standing would be no worse than any other such realism. Someone’s
disposition to find “moral facts” rather than “noble facts” explanatorily necessary owes to
the contingent facts of that person’s being constituted so as to find those facts explanatorily
necessary. “Since the normativity of any one [different possible realism] for any individual
rests if it exists at all in the desires and feelings she happens to have, nothing about that
value system makes it of its nature any more choiceworthy than any other.”

It turns out that the “realism” on offer from the synthetic naturalists leaves us mired in a severe arbitrariness problem. To state the matter plainly, there are no independent, authoritatively prescriptive moral reasons left to the deliberating agent by which one course of action is uniquely justified, since the competing and contradictory reasons that people on the other side of the issue believe justify their choices have exactly the same sort of grounding he believes his own reasons have. “But such arbitrariness,” Morgan rightly notes, “is seriously at odds with the moral perspective.” Morality naturally comes to us in terms of the obligation attending a course of action that is uniquely justified by independent, authoritatively prescriptive moral reasons that the moral agent recognizes within his deliberations.

Of course the “synthetic realists” (if Morgan is right about them, as I think he is) are far from the only philosophers who lead us into the problem of moral reasons arbitrariness. We have already seen how Gilbert Harman and also how sociobiologists like Ruse plunge headlong into this sort of arbitrariness. It also appears in Richard Rorty’s work, at those places where he speaks of “ethnocentrism,” “solidarity,” “we-intentions” and such. When Rorty insists that “we have to start from where we are… we are under no obligations other

---

183 Morgan, “Naturalism and Normativity,” 333.

184 This is one of the “objective pretentions” of morality that Allan Gibbard argues any moral philosophy must not fail to account for in some fashion on pain of losing a great deal of plausibility. See Chapter 3, section 1B below.

that the ‘we-intentions’ of the communities with which we identify,”\(^\text{186}\) he leaves us with no way to understand “our” reasons for action as uniquely justified in comparison to those of others. Just as some Muslims believe they have reasons that justify female circumcision because they are in the community they are in, so it is with “our” reasons when “we” take ourselves to be justified in opposing the practice.\(^\text{187}\) Moreover, there is an uncanny similarity between these statements by Rorty and those made by Simon Blackburn as he unpacks his non-cognitivist “quasi-realism.” Blackburn states that the belief of some fanatic Muslims that women who talk to men outside of their family should be killed “is false, pure and simple,” and “the fact that some people in Pakistan believe it and act on it is neither here nor there.”\(^\text{188}\) That sounds promising. But then he continues: “In saying this I am, of course, expressing my own attitude on this matter, and I hope that of many others. But that is nothing to be ashamed about. It is my attitude, and indeed I am a Western liberal with a strong sense of sympathy for the downtrodden. And that is the way to be, and if you are not like that then you can expect me to be in conflict with you.”\(^\text{189}\) But it is difficult to see how such proud, ringing declarations justify anything—especially when we realize that “we” can expect “them” to reply to “us” in kind. Although I will not argue the point here, it is


not clear that Blackburn has advanced much, if at all, over Rorty with respect to moral reasons arbitrariness. 190 “We” happen to have our attitudes and “they” happen to have theirs. What “they” take to be the authoritatively prescriptive moral reasons justifying their actions emerge from the same non-moral forces as “our” reasons, and what sets apart the reasons “we” accept is the philosophically rather uninteresting fact that they are “ours.”

There are many more examples that could be highlighted, 191 but these should be sufficient to show that problem of moral reasons arbitrariness is in fact common in moral philosophy.

Summarizing all this discussion, we can say that if a moral philosophy fails to establish support relations for moral judgments that uniquely justify those judgments in terms that make essential reference to a person’s ability to consider and weigh those support relations in making a decision about what should be done in a particular situation, that view suffers from moral reasons arbitrariness. This can happen when a moral philosophy explains moral reasons as resulting from non-moral forces that are themselves unresponsive to moral considerations (as when what I take to be “morally justified” owes merely to an “implicit agreement” with a group of people whose cooperation I happen to need because I was born in Afghanistan). It can also happen when a moral philosophy establishes justificatory grounds that may be shared by people making conflicting moral judgments, so that no particular moral judgment is uniquely justified amid the welter of conflicting moral judgments, all of which appeal to the same justificatory ground (as seen in the case of the

---

190 Seiriol Morgan suggests the same thing in “Naturalism and Normativity,” 344. See also Russ Schafer-Landau’s critique of non-cognitivism in Moral Realism, 22ff.

191 Elsewhere I have argued that Jürgen Habermas and Richard Brandt both succumb to what is, in effect, the problem of moral reasons arbitrariness, although I did not use this terminology at the time. See Bradley N. Seeman, “Whose Rationality? Which Cognitive Psychotherapy?” and Seeman, “Peirce, Habermas, and Moral Absolutes.”
“synthetic realists”). Perhaps there are other ways in which moral reasons arbitrariness can infect a moral philosophy, but I take no position on that here. In any case, when a person believes such explanations they cannot be isolated from the realm of moral deliberations because the very fact of this belief makes them a consideration to be weighed alongside other considerations.

This last point bears on one objection that should be considered briefly. Everything that has been argued above about moral reasons arbitrariness may be moot if an objection arising from Thomas Nagel’s work has no reply. In lockstep with today’s philosophical orthodoxy, Nagel declares: “I take it for granted that the objectivity of moral reasoning does not depend on its having an external reference. There is no moral analogue of the external world.” 192 Given this starting point and his views of human beings, I think Nagel’s moral philosophy will succumb to moral reasons arbitrariness—indeed, for reasons very much like the reasons Christine Korsgaard’s philosophy succumbs to the problem (as I will argue in Chapter 4). Nagel, however, would object right at the start: “someone who abandons or qualifies his basic methods of moral reasoning on historical or anthropological grounds alone is nearly as irrational as someone who abandons a mathematical belief on other than mathematical grounds.” 193 The point is that just as only mathematical reasoning is relevant to mathematical beliefs, so only moral reasoning is relevant to moral beliefs. In the same way we need good mathematical grounds to give up mathematical beliefs, we need good moral grounds to give up fundamental moral principles. Thus it makes no sense to give up a belief


in the value of equality just because someone points out that you would have held a
different belief if you had grown up in a caste society rather than in a culture deeply
influenced by Christian ideals. As Nagel puts it, in a passage worth quoting at length:

Suppose someone says, for example, “You only believe in equal opportunity because
you are a product of Western liberal society. If you had been brought up in a caste
society or one in which the possibilities for men and women were radically unequal,
you wouldn’t have the moral convictions you have or accept as persuasive the moral
arguments you now accept.” The second, hypothetical sentence is probably true,
but what about the first—specifically the “only”? In general, the fact that I wouldn’t
believe something if I hadn’t learned it proves nothing about the status of the belief
or its grounds. It may be impossible to explain the learning without invoking the
content of the belief itself, and the reasons for its truth, and it may be clear that what
I have learned is such that even if I hadn’t learned it, it would still be true. 194

We are pushed toward universal, moral questions, and these questions “require an answer of
the appropriate kind… They cannot be ruled out of order by pointing to something more
fundamental—psychological, cultural, or biological—that brings the request for justification
to an end. Only a justification can bring the request for justifications to an end.” 195 Many of
the worries about moral reasons arbitrariness are thus misplaced, according to Nagel. In
effect, the worries simply don’t trade in the language in which the question must be asked:
the normative language of universalized questions and justificatory deliberation about them.

Now, as will emerge in the Conclusion, I have a great deal of sympathy for a lot of
what Nagel says, especially in regards to “the independent force of the first-order [moral]
judgments themselves.” 196 Indeed, I believe some elements of those judgments have
enough force to resist a lot of bad theory. But, as an objection to moral reasons

194 Nagel, The Last Word, 103.
195 Nagel, The Last Word, 106.
196 Nagel, The Last Word, 103, passim.
arbitrariness, Nagel’s arguments misfire. First, when a person believes an explanation of moral reasons offered by a moral philosophy, taking it to be actually the case (as Nagel rightly sees we all do with whatever point of view we accept), that fact of belief introduces the explanation into the deliberative realm where moral judgments are forged by weighing the support for and against a particular judgment. It becomes a consideration, with the ability to support some positions and remove support from others. The explanatory fact, believed to be the case, can change the way other considerations are understood and how they will enter into one’s deliberations. The explanation does not thereby become a directly moral consideration, but it does become a rational consideration that factors into how things fit together in one’s overall view of reality and how it makes sense to act within that reality.

Pace Nagel, as one deliberates about what to do from within that overarching view of “the way the world is” (in Nagel’s own phrase), if the explanatory account shaping that view enshrines ontological categories that leave the person believing herself to inhabit a world

\[197\] Would this mean that if an evolutionary explanation along the lines of a “moral brain” were shown to be true, that this would underwrite justificatory considerations in their own terms (in contradiction to some things I say below)? I don’t think so, but a full answer would require much more space than I have. In short, while it would become a consideration, the nature of that consideration still matters. And in this case the non-moral form of the explanation still serves to undermine the moral reasons. The consideration, in other words, is still to the effect that my moral reasons are in the grip of something non-moral.

Indeed, I am tempted to go farther and say that when non-moral explanation is taken to be ontologically basic, in the sense that what it says is the case was before the justificatory realm came to be (as I take it that materialism as I understand it in the Introduction must hold), then it in some sense envelopes the justificatory realm and justification cannot operate autonomously without keeping an eye on what makes sense as fitting into that ontologically fundamental, non-moral explanatory story. As Harman puts it, “consider what it is for someone to have a sufficient reason to do something. Naturalism requires that this should be explained in terms congenial to science. We cannot simply treat this as irreducibly normative,” (Harman, Explaining Value, 86; emphasis added). After all, the stuff materialist science will countenance was before persons or deliberations or justifications came to be. Harman, as I see it, rightly refuses to grant ethics any independence from materialist explanation; for a materialist, somehow the non-moral, material stuff must have given rise to whatever else there is—including whatever we are pleased to call morality.

On the other hand, if persons and deliberations and justifications were first, then all explanation would be encompassed by these. Perhaps Love, and any demands that flow from it, were before all else—and any explanation would have to show how it fit into those purposes. This sounds weird to today’s philosophical impulses, but maybe it’s true for all that.

But all of that is tied to bigger questions which I have foresworn here.
where the support she seems to see for her moral judgments actually does arise from non-moral factors or from an equal-opportunity justificatory structure, then she will naturally consider whether the “reasons” that at first blush seem to support a moral position do in fact carry the weight she thought they did.

A second reason why Nagel’s argument in *The Last Word* does not dissolve the problem of moral reasons arbitrariness is that there is a disanalogy between math and morals. Simply put, morality does not enjoy the undisputed universality and constancy of mathematics, and this invites serious skeptical accounts in ways mathematics does not. Most people would find a skeptical account that invited them to rethink their notion of addition as just plain silly, along with Nagel. And there would be no examples of different ways of adding things to point to. Not so with morality. Now clearly we can overstate the case here, and the reasons for the variations in moral judgments are open to debate (maybe people forsake rational reflection, maybe they rebel, etc.); but the fact remains, there are people who don’t believe in Rawlsian “justice as fairness” or that all humans have a claim not to be dismembered. This simply does give worries about moral reasons arbitrariness more initial traction than worries about mathematical reasons arbitrariness could ever have. The challenge of moral reasons arbitrariness is more severe than Nagel allows.

Finally, at most Nagel’s arguments point out that the problem of moral reasons arbitrariness does not foreclose the possibility of an additional story that could be told that shows that moral reasons do survive the challenge of moral reasons arbitrariness. As Nagel himself puts it in the lengthy passage quoted above, “it may be impossible to explain the learning without invoking the content of the belief itself, and the reasons for its truth, and it
may be clear that what I have learned is such that even if I hadn’t learned it, it would still be true,” (emphasis added). Nagel rightly points out that the mere statement of moral reasons arbitrariness does not rule out of account one or another possible additional justificatory story, and these should be considered (as three will be here). But this justificatory story is owed, and it must be shown how it meets the challenge of moral reasons arbitrariness. Moreover, Nagel’s arguments clearly don’t preclude the possibility that something about materialist or other explanatory stories might undermine moral reasons generally. Though I believe such an argument can be made with respect to materialism, this dissertation does not take up that burden. I believe, however, that the problems exposed (as I hope) below in the justificatory stories put forward by Gibbard, Korsgaard, and Post, along with Seiriol Morgan’s argument against the “synthetic realists” as canvassed briefly above and Harman’s forthright admissions about where he takes materialism to lead in the moral realm should give one pause to think.

The upshot of all this is that Nagel’s arguments do not blunt the problem of moral reasons arbitrariness. Indeed, mutatis mutandis, his own moral philosophy can be seen to succumb to the problem in the same basic ways Christine Korsgaard’s moral philosophy does. Moral reasons arbitrariness is a stiff challenge to contemporary moral philosophy. Moreover, I agree with Seiriol Morgan that the ramifications of the kind of arbitrariness in question are extensive: talk of “discovering” what is valuable or right rings hollow, we are left without moral guidance, moral criticism is reduced to preference for the explanations one happens to find necessary due to one’s own contingent constitution, and moral
obligation loses its objective bite. 198 Within the shell of morality left to us, “there are no facts beyond the facts about what we happen to desire that rationally oblige us to make judgments of one kind rather than another.” 199 These and other problems with moral reasons arbitrariness were already touched on in the Introduction, and I won’t repeat all of them here. From all of these problems, however, I take it to be clear that if some moral philosophies suffer from moral reasons arbitrariness they should be rejected, and we should make a serious effort to develop an alternative approach to moral philosophy in hopes that a way of avoiding moral reasons arbitrariness may be found.

IV. Chapter Summary and a Look Ahead

In this chapter we began by looking at how deeply notions of independent, authoritative prescriptivity are embedded in our normal ways of talking about moral reasons, and I argued that these notions can be ripped out of our thought about reasons only at great cost. From there the focus shifted to questions of what issues can be clearly isolated from out of the tangled mass of internalism/externalism verbiage. Drawing on the work of William Frankena and Russ Shafer-Landau, I attempted to sketch the topography of internalism and externalism and to place my own position within this sketch. The major landmarks, I argued, have to do with the divide between reasons internalism/externalism and judgment internalism/externalism. With regard to the former issue, I identified myself as a


199 Morgan, “Naturalism and Normativity,” 337. See also the points raised by Shafer-Landau, Moral Realism, 22ff.
reasons externalist, while on the latter, I gave reasons for favoring judgment internalism, and I gave some reason for thinking that these positions fit together nicely.

After this close look at moral reasons, we made our initial approach to the problem of moral reasons arbitrariness by considering Christine Korsgaard’s worry about the “Why be moral?” question when this is taken as a question about the justification of morality, especially when the going gets tough. This worry I defended against Harman and his attempts to collapse the distinction between justifying and motivating reasons. Though I didn’t need to in order to establish the point this dissertation is attempting to make, I argued that Harman’s objections failed to undermine the distinction, especially in the face of such strong testimony from the moral experience of at least many (perhaps all) people. In the end, Harman’s moral philosophy was seen to illustrate the problem of moral reasons arbitrariness, and I suggested that Harman’s arguments may be right, so long as one accepts his materialist starting point. I also introduced two test cases that will feature prominently in the following chapters: (1) the case of the contented criminal, and (2) the case of the viable dictator.

Finally, I have just finished arguing that there is a real problem of moral reasons arbitrariness, and I have suggested that it is widespread in moral philosophy. The problem, to sum it up, is that when the independent, authoritatively prescriptive moral reasons (what I call “external moral reasons”) about which moral agents deliberate and to which they respond are thought by that agent to be in the grip of non-moral forces that similarly determine what other moral agents take to be the reasons justifying their (opposing) actions, then that moral agent is stripped of reasons for taking her own actions to be uniquely
justified. Both in this chapter and in the Introduction, I identified several ways in which such moral reasons arbitrariness undermines the coherence of moral deliberation.

That the problem of moral reasons arbitrariness has infected moral philosophy there can be no doubt. But how widespread is it? Although I will not be arguing the point in this dissertation, I believe that the problem is seldom avoided in moral philosophy today, owing to the metaphysical starting point assumed by most contemporary moral philosophers. This starting point is materialism, and it is clear that in a stark and undigested form such as that advanced by philosophers like Gilbert Harman and Michael Ruse, this metaphysical starting point does lead to moral reasons arbitrariness. The real question, as I intimated above, is whether some additional story can be told that shows how authoritatively prescriptive moral reasons fits into a materialist metaphysic, or that somehow rehabilitates those reasons.²⁰⁰ Now I do believe that there are some general flaws fatal to these additional stories, but the remainder of the dissertation will be content to identify the problems with the stories Allan Gibbard, Christine Korsgaard, and John Post each add to the materialist story to try to show that materialist explanations of human moral behavior don’t undermine the authority of the reasons we take ourselves to be responding to.

I shall argue that, at least in the case of Gibbard, Korsgaard, and Post, the attempts fail. Each of these moral philosophers offers an account of moral reasons that is not robust enough to avoid the problem of moral reasons arbitrariness, failing to capture the independent, authoritatively prescriptivity characteristic of moral reasons. It is to that

argument that we now turn, examining Gibbard’s attempts to vouchsafe moral reasons that can support a morality worthy of the name.
CHAPTER THREE
GIBBARD ON RATIONALITY, PLANS, AND THE
“OBJECTIVE PRETENSIONS” OF NORMATIVITY

Upon realizing that Allan Gibbard’s “norm expressivism” revolves around a noncognitivist account of rationality, one might be puzzled. Noncognitivism does not immediately seem promising as an account of rationality. As Nicholas Sturgeon notes, “noncognitivism about epistemology is a lot to swallow, and even Gibbard forgets that he is committed to it.” It is difficult indeed to maintain such a stance consistently, and yet Gibbard’s commitment to a noncognitivism about rationality runs deep; for it is from rationality, construed in this way, that he seeks to give a materialist account of the normativity central to human moral experience. Gibbard ties normativity generally and morality specifically to a noncognitivist account of rationality. Gibbard’s noncognitivist moral philosophy pivots on his noncognitivist account of rationality.

Perhaps one would expect Gibbard, given such a starting point, to dismiss the phenomenological surface of the human moral life, where there is at least a claim to authority and objectivity; but Gibbard clearly sees that “any account… that ignores this claim must be

---

1 Nicholas L. Sturgeon, “Critical Study: Allan Gibbard, Wise Choices, Apt Feelings,” Noûs 29, no. 3 (1995): 403. Gibbard’s noncognitivism about rationality is seen in statements to the effect that we “have no sharp notion of cognitive judgment to apply” (Allan Gibbard, Wise Choices, Apt Feelings: A Theory of Normative Judgment (Cambridge, MA: Harvard University Press, 1990), 131). Indeed, Gibbard reduces even epistemic norms to matters of acceptance/endorsement. As he sees, a materialist is allowed no room for normativity as such, and he is determined to be consistent (curiously) and explain all normativity in terms permitted by his monistic ontology.

2 Note 17 in the Introduction gives my reasons for using the term “materialist” instead of “naturalist.” In close proximity to quotations where Gibbard uses the word “naturalism” I sometimes relax this practice.
defective.” Gibbard thus sets himself a task: “My hope, then, is to save what is clear in ordinary thought about rationality,” which would, in turn, preserve whatever is needed of the “objective pretensions” of our normative life generally. Yet at every turn Gibbard insists that this phenomenological surface of human normative activity be seamlessly integrated into what he calls the “Galilean” ontology of nothing but material stuff that gives rise to everything else. “Gibbard’s grand strategy is to make do with a brass-tacks Galilean core of descriptive language and an all-purpose normative term, roughly, ‘is rational’.” Using this one technical term, Gibbard believes he can salvage much of what he calls the “objective pretensions” of human moral life, while at the same time avoiding anything fishy from a materialist standpoint. The trick, for Gibbard, is to pack everything about normativity that a materialist can hope to salvage into this one notion, “rationality,” where that is understood not as a faculty that gets hold of a property—“the rational”—but as a human activity wholly comprehensible in “Galilean” terms. An explanatory account of human normative activity “should cohere with our best naturalistic accounts of normative

3 Gibbard, *Wise Choices, Apt Feelings*, 153. As Gibbard notes elsewhere, “As we lead our normative lives we need a sense of what we are doing. A picture of ourselves that can guide us” (8).


5 Gibbard, *Wise Choices, Apt Feelings*, 120-25. It is far from clear that Galileo himself would have been happy to find this materialist metaphysical outlook named after himself, but this chapter will follow Gibbard’s usage. Gibbard’s use of this term is slippery in that it allows some wiggle room for the materialist to accept mental properties. That materialists can actually account for the mental in its own terms is, of course, highly controversial, but most materialists want to allow for some sort of “property dualism.” Jaegwon Kim is (rightly) very critical of such attempts to have one’s materialism “on the cheap” (as he puts it). See Jaegwon Kim, *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation* (Cambridge, MA: MIT Press, 1998), 120. These questions have been treated in Bradley N. Seeman, “Out-Kimming Kim,” but a further exploration of them would go beyond the scope of the present thesis.

In attempting to meet these demands, Gibbard offers an ingenious and subtle set of reflections about a wide variety of important problems in moral philosophy, with his answers leaning heavily on his speculations about how evolution might have worked at key junctures of this story.

The present chapter focuses attention on the linchpin of Gibbard’s account of normativity, namely, his attempt to leverage a notion of “rationality” to reconcile the “objective pretensions” of the human normative life with the Galilean world of a materialist science. In what follows, I argue that Gibbard’s attempt to effect this reconciliation fails due to the inability of the kind of moral reasons vouchsafed by his notion of rationality to meet the challenge of moral reasons arbitrariness. Since Gibbard advocates “starting out with questions of rationality, and interpreting moral questions as special questions of rationality,” his ability to salvage something significant from the objective pretensions of human moral phenomenology hinges upon questions of human rationality and the reasons it gives us. Indeed, in his most recent work, Gibbard goes so far as to argue that a moment of acceptance or endorsement can even ground external reasons. After sketching this account, Gibbard states that it “accords sense to reason claims that are ‘external.’ Anyone who accepts the account is thus an ‘external reasons theorist’… she maintains that there can be a reason to do something even when one’s ‘motivational set’ is devoid of anything relevant.”

Gibbard’s reasons for defending reasons externalism are sound, but his attempt fails to make the crucial transition from acceptance to external reasons, succumbing to problems of

---


8 Gibbard, Wise Choices, Apt Feelings, viii

moral reasons arbitrariness and thus failing to salvage much of interest from the “objective pretensions” of morality.

This argument will be pursued in three sections. The first section’s critique of Allan Gibbard’s *Wise Choices, Apt Feelings* will be the central focus of this chapter since it is Gibbard’s most important statement of his norm expressivism and all his subsequent work in moral philosophy builds on it, with subsequent departures being fairly minor.\(^\text{10}\) The first section begins by laying out the central themes of Gibbard’s norm expressivism, showing how it eschews descriptivism and attempts to hold together a “Galilean” metaphysics and the objectivity and authority central to the phenomenology of human normative activity. Gibbard identifies three central “objective pretensions” of morality—the pretense of independence, the pretense of non-arbitrary reasons, and the pretense of interpersonal authority—and particular attention will be paid to Gibbard’s attempts to salvage something substantive from these “pretensions” that will be acceptable to materialism. This leads into an exploration of problems that Gibbard’s non-cognitivist notion of rationality as a kind of acceptance creates for reasons; reasons, it turns out, simply express acceptance and so collapse into our activity of accepting what we will. Reasons in Gibbard’s system provide no independent constraint on the human activity of accepting one thing or another. From there it will be argued that whatever way Gibbard turns to salvage some sort of objectivity for morality, he fails to find a way forward. Instead, Gibbard runs headlong into the problem of moral reasons arbitrariness and does not make good his stated aim of reconciling “Galilean” metaphysics with the central elements of “ordinary thought” about

our normativity in day-to-day life. Gibbard, it turns out, cannot have it both ways—or so it shall be argued.

The second section examines Gibbard’s modifications to norm expressivism as laid out in his recent book, *Thinking How to Live*. Moving more quickly in this section, it will be argued that the modifications Gibbard makes in this work fail to solve the severe difficulties of his original statement of the position. Specifically, Gibbard places the notion of a “plan” at the center of his restatement of his expressivist account of normativity, focusing attention on the slogan “thinking what one ought to do is thinking what to do.”

11 Deciding what to do by committing to a plan thereby brings normativity into the world precisely through the very activity of planning to φ. Normativity arises from an activity rather than from a representation of what ought to be done that has cognitive content. As Gibbard puts it in his recent Tanner Lectures, “ought thoughts are like plans. Thinking what I ought to do amounts to thinking what to do.”

12 But this gambit will only succeed to the extent that Gibbard’s notion of a plan is able to shoulder the weight of normativity without either collapsing into normative reasons arbitrariness or else smuggling in cognitive or normative contraband. The second section of Chapter 3 argues that Gibbard’s notion of a “plan” faces the same insurmountable difficulties facing the idea of “rationality” at the heart of his earlier statement of expressivism: Reasons to plan to do one thing rather than another do

---


12 Gibbard, *Reconciling Our Aims*, 19; emphasis in original. Gibbard, under pressure from philosophers pushing the “embedding problem” (see note 45 below), takes great pains to show that there is a kind of representational content involved in these planning activities. The success or failure of Gibbard’s efforts to meet these technical challenges do not bear directly on the point argued here. The question pursued here is not whether the planning activity Gibbard is concerned with admits of some sort of construal that supports a kind of representation, but with whether these planning activities allow Gibbard to avoid moral reasons arbitrariness and fulfill his stated goal of capturing something of the “objective pretenses” central to moral phenomenology.
not precede the plan itself on Gibbard’s noncognitivism about reasons, and so nothing limits plans other than forces wholly unresponsive to normativity because there is only non-normative, materialist stuff in a “Galilean” metaphysic, and normativity must arise from that non-normative stuff.

The third and final section contrasts Gibbard’s stance on the role of the “second person” (the you or “thou” of our daily interactions) with Stephen Darwall’s fascinating work in his recent book, *The Second-Person Standpoint*. There it will be argued that Darwall’s notion of second-personal authority captures something crucial that has gone missing from Gibbard’s account, and that this insight should be pursued further as providing hope for the authoritatively prescriptive moral reasons addressed to all people that are needed to avoid the “coal pit” of moral reasons arbitrariness. In Chapter 4 I will argue that Darwall’s own account also fails—and for some of the same reasons that Gibbard’s moral philosophy is inadequate—but he shows us part of what is needed for authoritatively prescriptive moral reasons that address all people, namely an authoritative second-person address.

Darwall’s work thus serves as something of a foil to Gibbard’s project, showing the importance of second-personal reasons—reasons that go missing from Gibbard’s moral philosophy. But before considering this instructive contrast, there must be a clear understanding of Gibbard’s project and why it succumbs to moral reasons arbitrariness. It is that task that will now be taken up.
I. Rationality and Reasons in Allan Gibbard’s

*Wise Choices, Apt Feelings*

A. Gibbard’s Norm Expressivism

At the heart of Gibbard’s project of “norm expressivism” stands his noncognitivist account of rationality. “Rationality,” according to Gibbard, “figures in all aspects of human life and thought... using the term we can explain thoughts of morality, or what is worth having, or whether a claim merits credence, or what is shameful and what is cause for pride—and indeed all the other kinds of thought that philosophers call normative.”

Gibbard puts the point flatly: “All norms, we might say, are norms of rationality,” where “rationality” is analyzed as a non-cognitive, expressivist notion. Gibbard hopes to analyze the human activities behind normativity in terms of this one notion, allowing his expressivism to permeate all the other normative notions he analyzes. “Rationality” will then be explicated in strictly “Galilean” terms by recourse to a speculative evolutionary psychology that hypothesizes that the kinds of activity behind human normativity evolved under pressures that selected for the coordinating activity of language that fosters cooperative action—cooperation that has been immensely successful, in evolutionary terms. It is this speculative evolutionary psychology that provides Gibbard’s substantive account of rationality, his story about why we endorse what we do. Thus Gibbard’s norm-expressivism rests on evolved human propensities to endorse this or that system of norms. In this way,

---


normativity can be analyzed in terms amenable to the Galilean metaphysics Gibbard prefers.\textsuperscript{15}

Gibbard makes clear his belief that the kind of rationality compatible with his Galilean metaphysics is a noncognitivist rationality arising from the human activity of acceptance or endorsement. Gibbard takes pains to distance himself from descriptivist accounts of rationality, making this plain even in the term he adopts for his own view: “expressivism.” It is critically important to note two prongs to Gibbard’s critique of descriptivism. First, as Gibbard observes (rightly, it would seem) the Galilean world has no place for a special kind of property that makes something rational; there are no such facts “out there” among the fundamental entities countenanced in the Galilean world. “We cannot rest content with magic,” Gibbard urges, “and the simple picture of facts and minds that grasp them is magical.”\textsuperscript{16} Moving away from this “magical” picture, Gibbard states that “a non-cognitivist agrees that some judgments are distinctly normative, but not, he says, because they are judgments of normative fact. They are not judgments of fact at all, and there are no peculiarly normative facts.”\textsuperscript{17} Continuing on to refer to his “speculative evolutionary account,” Gibbard claims that “if the account is on the right track, then our normative capacities can be explained without supposing that there is a special kind of normative fact to which they typically respond.”\textsuperscript{18} Instead of properties that make something rational upon being somehow mysteriously grasped by our minds, it is human

\textsuperscript{15} “To say it ‘makes sense’ to have such-and-such an attitude is roughly to express one’s acceptance of norms that permit the attitude. Expressing one’s acceptance of norms is something we can understand in psychological terms, as a part of that natural world,” (Gibbard, Wise Choices, Apt Feelings, 248).

\textsuperscript{16} Gibbard, Wise Choices, Apt Feelings, 106.

\textsuperscript{17} Gibbard, Wise Choices, Apt Feelings, 107.

\textsuperscript{18} Gibbard, Wise Choices, Apt Feelings, 107.
activity that first brings rationality to a non-normative, materialist world. Like Hume, Gibbard holds that there are no normative facts in the natural world for human beings to somehow “grasp” and describe. As Gibbard declares, “there is no class of facts such that a judgment is normative if and only if it naturally represents a fact in that class.” Rationality, as Gibbard understands it, is not a matter of accurately describing something about the world, correctly ascribing some property to some part of that world.

The “magic” tainting some descriptivist accounts simply cannot be abided by those committed to the Galilean metaphysic. To borrow language from J.L. Mackie, the materialist may not countenance anything “queer” like “magical” rational-making facts residing somewhere. This first prong of Gibbard’s anti-descriptivism is important to Gibbard’s Galilean metaphysic; for if rationality were somehow “out there” in the world independently of human minds and their activity in some Platonic sense, this would obviously create problems for materialism. “On the Platonist picture, among the facts of the world are facts of what is rational and what is not.” If there were “rational-making” properties in the mind-independent world, as opposed to concepts of “the rational” bearing the

---

19 Gibbard’s fellow “quasi-realist” Simon Blackburn, explicitly references the Humean metaphor of “spreading” valuations onto naturalistic matters of fact in the title of the book that lays out the core of the quasi-realist program: *Spreading the Word* (Simon Blackburn, *Spreading the Word: Groundings in the Philosophy of Language* (Oxford: Clarendon Press, 1984). Gibbard frequently notes the similarity of his views to Blackburn’s views, and their shared expressivism is certainly basic to their common outlook on the place to start a materialist analysis of morality.

It should be noted that Gibbard is critical of Hume’s attempts (or the attempts of some of Hume’s interpreters) to give a “substantive” analysis of “instrumental rationality.” By contrast, Gibbard insists on a noncognitivist account of rationality. See Gibbard, *Wise Choices, Apt Feelings*, 9-18. I am not convinced that Gibbard needs to take as much issue with Hume himself as he does, but this is not the place to attempt to exegete Hume and I will not argue the point one way or the other.


stamp of human manufacture, then matters would get quite sticky for those who would explain everything in terms of a Galilean metaphysic.\(^\text{23}\) Simply put, if the world prior to the animating touch of human activity were to contain a special class of rational-making facts, that world would be quite inexplicable in strictly Galilean terms. “Nothing in a plausible, naturalistic picture of our place in the universe requires these non-natural facts and these powers of non-sensory apprehension.”\(^\text{24}\) In an indifferent, non-normative Galilean world, there is no question of recognizing or describing facts that are somehow normative in themselves. No special kind of rational-making facts in that world stake a normative claim on these beings. Gibbard’s “norm-expressivism is meant to capture whatever there is to ordinary notions of rationality if Platonism is excluded.”\(^\text{25}\)

Gibbard simply dismisses such “Platonism,” wanting only to “debunk” it.\(^\text{26}\)

While Gibbard does not seriously consider a “magical” type of descriptivism, he does carefully consider non-magical forms of descriptivism, if only to reject them in the end. The second prong of Gibbard’s anti-descriptivism undermines these materialistically kosher

\[^{23}\text{Gibbard sees the problem clearly. In dealing with a related issue of a human normative faculty (which Gibbard rejects), Gibbard gets at the central issue: “I should stress why I could not myself call the faculty of normative control ‘reason’, or use the word ‘rational’ to mean ‘dictated by the faculty of normative control’. I have avoided these terms because they are normative as well as descriptive. If I were to speak of ‘rational control’, I would be hard put to avoid the suggestion that what is done under ‘rational control’ is rational in the normative sense –that it is something that it really makes sense to do” (Gibbard, Wise Choices, Apt Feelings, 80). It is critically important for Gibbard’s project that there be nothing that it “really makes sense to do,” in the sense that it is somehow prescribed by a mind-independent reality. Such a reality has no place in the Galilean world. No such properties can exist, either within the world independent of our activity or in some faculty that really does track that which is to be done.}\]

\[^{24}\text{Gibbard, Wise Choices, Apt Feelings, 154. It does not seem that Gibbard would be more favorably disposed to the varieties of theistic materialism that have begun to crop up. For an overview and critique of such views, see Charles Taliaferro and Stewart Goetz, “The Prospect of Christian Materialism,” Christian Scholar’s Review 37, no. 3 (Spring 2008): 303-21.}\]

\[^{25}\text{Gibbard, Wise Choices, Apt Feelings, 154.}\]

\[^{26}\text{Gibbard, Wise Choices, Apt Feelings, 154.}\]
forms of descriptivism. The descriptivist view Gibbard takes most seriously is that of his former teacher and colleague, Richard Brandt. Brandt advances a full-information account of rationality grounded in rational intrinsic desires where, “the aim is to show that some intrinsic desires and aversions would be present in some persons if relevant available information registered fully, that is, if the persons repeatedly represented to themselves, in an ideally vivid way, and at an appropriate time, the available information which is relevant in the sense that it would make a difference to desires and aversions if they thought of it.”

Gibbard believes that all descriptivist views, Brandt’s included, founder on an insuperable difficulty: “any such descriptivist analysis leaves a puzzle. It misses the chief point of calling something ‘rational’: the endorsement the term connotes.” As Gibbard sees it, descriptivist analyses of rationality always fail to capture the recommending force of that which is rational; at the conclusion of any such analysis what has gone missing is the sense that whatever the analysis has delivered is “right” or “ought to be done” or “of value.” Accordingly, after stating once again that the word rational “has an automatically recommending force,” Gibbard notes that “rationality in Brandt’s sense… need not be a recommendation.”

---

27 Brandt, *A Theory of the Good and the Right*, 111. See also, Richard B. Brandt, “Rational Desires,” in *Morality, Utilitarianism, and Rights* (Cambridge: Cambridge University Press, 1992). Brandt’s moral theory is founded on the idea that the “intrinsic desires that it would be rational for a person to have are those that she would have if she were to represent all available facts in an ideally vivid way,” (Thomas L. Carson, “Gibbard’s Conceptual Scheme for Moral Philosophy,” *Philosophy and Phenomenological Research* 52, no. 4 (December 1992): 953). Growing out of my work for Carson’s “History of Ethics” course, I have interacted seriously with Brandt’s full-information theory of rationality and the moral theory he develops in connection with it. See Seeman, “Whose Rationality? Which Cognitive Psychotherapy?”


Gibbard gives a number of examples of this problem in Brandt’s full-information analysis. Imagine a faithful civil servant who steadfastly refuses bribes, but who realizes that he would quite possibly change his mind if he were to be fully and vividly informed about all the pleasures that accepting bribes would allow him to experience. On Brandt’s account, the civil servant is acting irrationally in not taking steps to be fully and vividly informed and so likely starting to accept bribes. But, as Gibbard asks: “Does that recommend dwelling on the temptations? Does that recommend bribery? The honest civil servant might well think not.”

The point, of course, is that the thing that Brandt’s account recommends as “rational,” one might well think is something that many rational people—such as the civil servant—might not recommend or endorse. Indeed, they could understandably see it as irrational. Other cases like that of the civil servant are not difficult to formulate, and Gibbard advances several others that need not be considered here.

Gibbard holds that these “funny cases—the cases where Brandt’s account labels crazy acts rational—have a systematic import.” The import is that Brandt’s descriptivist analysis of rationality fails to capture “the element of endorsement that full-information accounts leave out,” and in this element “lies the specially normative aspect of that term.”

Revisions of the full-rationality account may be offered, but Gibbard will meet them all in the same way. Thomas Carson, for example, proposes that a full-information account might be saved by holding that “rational or ideally rational desires are those one would have had if one had been fully informed (and free of cognitive mistakes) at all times at which the desires

32 Gibbard, Wise Choices, Apt Feelings, 22.
were being formed.”33 Gibbard replies according to form: I might (possibly correctly) come to believe that my altruistic impulses rest on my having mistaken the cries of other babies as my own—a mistake all babies make. “Once convinced of this,” Gibbard asks, “would I conclude that all my intrinsic concern for other people is irrational? I would if Carson’s proposal captures what I mean by ‘rational’, but I’m convinced I wouldn’t.”34 Once again, Gibbard points to a recurring problem he sees as infecting all descriptivist accounts: the crucial element of endorsement or recommendation comes apart from a description of the facts.35 Whether rightly or wrongly, it is clear that Gibbard rejects full-information accounts of rationality and morality, and measuring the success or failure of Gibbard’s own project involves taking the measure of his noncognitivist expressivism.

Gibbard presses his critique of descriptivism by pressing a crucial question at this point: “What, though, of the special element that makes normative thought normative?” He continues on to answer: “There is such an element, I am claiming, and it involves a kind of endorsement—an endorsement that any descriptivist analysis treats inadequately. The problem is not merely that every time one loophole in the analysis is closed, others remain. It is that a single loophole remains unpluggable by descriptivistic analysis”36 Accordingly,

33 Carson, “Gibbard’s Conceptual Scheme for Moral Philosophy,” 954.
35 Other counterexamples to a modified full-information account may lurk in the area, but pursuing these issues would lead far afield, given that it is Gibbard’s own views that are being examined in light of the threat of moral reasons arbitrariness. Perhaps a full-information account might be bolstered by considering how God might affect the analysis, but Gibbard will have no patience with that at all.
36 Gibbard, Wise Choices, Apt Feelings, 33. Although the question will not be explored here, one might ask if Gibbard’s own analysis is finally a descriptive analysis itself, insisting that normativity has to be analyzable in terms amenable to a materialist description of the world. Insofar as Gibbard’s own analysis even talks about the right thing, it ends up packing the strangeness of normativity into the particular primates who somehow do the endorsing. But simply trundling the strangeness from properties in the world over to a being
Gibbard insists that we humans bring normativity to an indifferent Galilean world by our activity of acceptance or endorsement; that which we accept or endorse is rational. This account will be nuanced some, but acceptance remains the core notion throughout. “To call a thing rational is to endorse it in some way.”

This leads into the heart of Gibbard’s strategy: “Instead of trying to define a property ‘rationality’ by giving conditions under which a thing would have that property or lack it, start with the use of the term. Fix on the dictum ‘To call a thing rational is to endorse it,’ and search for a sense of “endorse” for which the dictum holds true.” Thus, with this term “rational” Gibbard intends to capture “a kind of direct and flavorless endorsement,” and this renders rationality noncognitive at heart, being a matter of acceptance rather than getting facts right. It is of the essence of Gibbard’s staunch noncognitivism that what is “rational” arises from a human activity of endorsement or acceptance.

Gibbard recognizes that this is not the meaning the word “rational” usually has, but he does not seek to capture ordinary usage, hoping rather to devise a term of art that

---

37 Gibbard, *Wise Choices, Apt Feelings*, 6. Other statements building on this basic notion may be found on pp. 6-10, 33, 43, 45-6, 49, 81, 83, 91-2, 153, 162, 166, 189, 248, and 313; this idea is central throughout Gibbard’s book.

38 Gibbard, *Wise Choices, Apt Feelings*, 6. Gibbard means his use of double and single quotation marks to be significant here. The double quotations around “endorse” in the passage mean that he is using it as a technical term with his own, proprietary meaning.

signifies a sort of generic acceptance. For this purpose he lights on the term “rational.” Gibbard makes his intentions clear: “My real claim is not for the word ‘rational’, but for a meaning I want to exploit.” The meaning exploited has to do with a colorless, generic endorsement that is the essence of the human activity that gives rise to the normative and ramifies throughout the various flavors of human normative activity, encompassing not only rationality, but aesthetics and morality and, indeed, everything that is normative. “Rationality” is the crucial link, the place where the human activity of endorsement spreads itself onto the Galilean world in a way Gibbard believes can be accounted for in terms of a psychology explicable in strictly materialistic terms.

In summary, Gibbard attempts to pack all that passes for “normativity” among humans into this one term, “rational,” and to explicate this one term by means of another, namely “endorsement”—a term he hopes can be traced back to a human activity that is (1) evolutionarily explicable, so as to be acceptable to materialist sensibilities, and yet also (2) robust enough in terms of what it adds to a materialist account to allow for a reasonable facsimile of the “normativity” central to human experience. If the human activity of endorsement fails to bear the weight it needs to bear in accounting for normativity, then his materialist account ends in failure. Gibbard must show that normativity is “rationality” (in

40 One wonders here if there is any good reason not to stick with the words “accept” or “endorse,” which would seem to be the most straightforward. One might be forgiven for wondering if perhaps a main reason is that one would thereby lose a great deal of rhetorical flexibility and hortatory power.


42 Gibbard puts it this way: “I should stress why I could not myself call the faculty of normative control ‘reason’, or use the words ‘rational’ to mean ‘dictated by the faculty of normative control’. I have avoided these terms because they are normative as well as descriptive. If I were to speak of ‘rational control’, I would be hard put to avoid the suggestion that what is done under ‘rational control’ is rational in the normative sense—that it is something that it really makes sense to do,” (Gibbard, *Wise Choices, Apt Feelings*, 80).
his proprietary sense), and that rationality is a sort of endorsement explicable in materialist
terms.

But what is being accepted or endorsed in this picture, what sort of being does the
accepting, and what is it for that being to endorse whatever it is that is endorsed? Because of his
materialism, Gibbard cannot allow this normativity to reside in the world itself, and so he
insists that “to call something rational is not, in the strict sense, to attribute a property to it.
It is to do something else: to express a state of mind. It is… to express one’s acceptance of
norms that permit the thing in question.” In other words, the human normative language
orbiting the term “rational” does not make claims about facts of any sort, but rather expresses
acceptance of norms. Gibbard argues that what we accept are norms, or, in terms of
Gibbard’s attempts to develop a noncognitivist logic, we (ideally) accept “factual-normative
(FN) worlds.” In short, due to various contingencies we are thrown into a situation where

As a materialist, Gibbard attempts to avoid not only any normativity in the world apart from human beings, but also an inherent normative capacities in human beings themselves.


44 Had Gibbard taken a cognitivist turn here, it seems likely that he would have pursued an “error theory” of morality along the lines pursued by J.L. Mackie.

45 One of the thorny problems plaguing noncognitivist accounts like Gibbard’s is the Frege/Geach or “embedding” problem, and the apparatus of FN worlds is developed under pressure to respond to this difficulty. The problem is a hard one. It is less than clear what one attitude is being expressed in a sentence like “either John should lie, or else Christine should not go to the store.” This seems not to be expressing any single attitude, but to be doing something altogether different. What is the attitude being expressed here? There is rather a logical relation being stated here. But a logical relation of what? Expressed attitudes? Does it make sense to say “Either Go Cubsies! or else Yum, pizza!” That is puzzling. What attitude is being expressed here? The problem is (1) there seems to be no expression at all when expressed attitudes are run together, and (2) what is being done is a matter of logic that does not fit with expressions of attitudes at all. As Simon Blackburn puts it, when I say something like “Go Cubsies!” this is “not in the space of logic at all” (Simon Blackburn, “Gibbard on Normative Logic,” Philosophy and Phenomenological Research 52, no. 4 (December 1992): 947). The trick, as Blackburn points out, is to transform an expressed attitude, which is the wrong sort of thing to be negated and such, into “negatable content” (949), without drawing on any sort of cognitive content (such as consistency) in effecting the transformation (948). I would argue that Gibbard’s attempts pull off this nice little trick end in failure (as do Blackburn’s best efforts), but these problems will not be pursued further here.
we find the prescriptions of a particular group of people relevant for us since we need the cooperation of these particular people if a wide variety of the projects that interest us are to have any hope of success.\textsuperscript{46} It is in reference to our attempts to secure the cooperation of such groups that Gibbard defines a norm as “a possible rule or prescription, expressible by an imperative.”\textsuperscript{47} These imperatives come to expression through a community—both in its language and in its social training—and what an individual accepts are the imperatives of a group whose cooperation she needs.\textsuperscript{48} This acceptance provides the raw material for a FN world.

Cooperation is thus the leverage point for Gibbard’s evolutionary story about how we come to accept norms, and he expends a great deal of effort to develop a plausible empirical psychology that focuses on our need to get along with others. “Expressing one’s acceptance of norms is something we can understand in psychological terms, as a part of the natural world.”\textsuperscript{49} Human moral psychology does not access normative facts about the world, but rather expresses the results of materialistically explicable psychological processes. Here Gibbard leaves the realm of philosophical analysis of concepts and formulates a psychology informed by evolutionary speculations about what function human normative

\textsuperscript{46} Although this dissertation is not the place to pursue the idea, it would be interesting to explore Heidegger’s notion of “thrownness” (later developed by Sartre) in this connection.

\textsuperscript{47} Gibbard, \textit{Wise Choices, Apt Feelings}, 46; elsewhere Gibbard holds that a “norm is simply an imperative” (70). At this point we can see the legacy of earlier noncognitivists, like Hare, \textit{The Language of Morals}.

\textsuperscript{48} Gibbard understands language as playing a central role in the coordination of human behavior. “What is it, then, to accept a norm? To understand acceptance we should look to language” (Gibbard, \textit{Wise Choices, Apt Feelings}, 71). Willingness to avow a norm in discussion with others, along with a willingness to act consistently, constitutes acceptance of a particular norm (73-5).

\textsuperscript{49} Gibbard, \textit{Wise Choices, Apt Feelings}, 248. Or again, “I myself need to look at human psychology… an expressivist analysis like mine requires it. The analysis reduces the meaning of normative terms to psychology” (25-6).
activities might have served that allowed for their selection. “If there is such a thing as governance by norms, there must be psychic mechanisms that accomplish it, and we can ask about their biological function. That function, I want to suggest, is to coordinate.”

The kind of normative behavior we humans exhibit is driven by the benefits we reap when our behavior is coordinated in ways that facilitate our working together on various projects that allowed us to survive better than competitors who did not evolve the mechanisms needed to coordinate behavior in complex ways.

To make a long story short, Gibbard thinks that accepting the norms of a relevant group of other human beings tends to deflect or defuse their anger and enlist their cooperation, thus increasing our ability to work with them. One such behavior is a willingness to stand within a system of mutual demands—a system vital to normative consensus and successful cooperative action. “A person who refuses these demands must therefore be a poor candidate for cooperation of any kind—and in human life, cooperation is vital. It is fitness-enhancing, then, to stand ready to engage in normative discussion, and so to accept the demands for consistency that are part of the package.”

Selection pressures have led to the human drive to coordinate our actions with others, with one central development being language and the drive to enter into normative discussion with relevant others and the rules of reciprocity that tend to govern it. There are pressures toward consistency and away from the “strain” of continually monitoring a “double life” that tends

---


51 “A person invariably depends on intricate systems of cooperation and reciprocity if he is to have any decent chance of survival, reproduction, and the fostering of his children” (Gibbard, *Wise Choices, Apt Feelings*, 138).

to bring our self-governing norms into line with the norms we find it useful to avow in
discussion with others whose cooperation we need.

The same forces were at work in the development of fitness-enhancing emotions.
Instinctual behaviors that damaged cooperation by engendering conflict with people needed
for one’s projects were pressured out of the gene pool. Thus the psychic mechanisms
guiding human moral behavior “will be shaped to respond, in a rough way, to clues about
the kinds of strategies that pay in various different kinds of circumstances—pay
reproductively.”

At the same time, behaviors that led people to avoid creating anger and
to take actions to mollify powerful or otherwise relevant others when one had angered them
were selected (one’s projects being thus furthered, leading to better reproductive chances
and better success in rearing young). Human beings thus became sensitive to certain
emotions—particularly guilt and shame—that “typically portend bad treatment from
others.”

“One’s chances of damaging conflict are reduced, then, if one feels guilty when
guilt and its normal accompaniments are demanded by others, and if one demands guilt and
its normal accompaniments only when others are prepared to feel guilty. Hence it tends to
be advantageous for an individual to coordinate his guilt with the resentment of others and
his resentment with the guilt of others.”

Gibbard states flatly, “I identify moral concepts
by their tie to particular moral sentiments: To call an act morally ‘wrong’ is, very roughly, to
say that it warrants guilt and anger.” In this way Gibbard believes he can explain the

feelings that drive our normative life, and he resists attempts to broaden the moral emotions beyond guilt and anger because only these capture moral wrongness—a sense of having violated a binding norm. When Gibbard is pressed to recognize a broader emotional base for the evolutionary moral psychology he seeks to develop, he insists that while this would technically be compatible with his expressivism, “the question is whether the upshot would still be morality, and whether any of the concepts we would then be using is our familiar concept of moral wrongness.” In short, the guilt human beings feel upon violating an obligation has arisen from evolutionary pressures as a way of signaling a looming and costly conflict with others. Thus at the heart of Gibbard’s account we find the evolutionary benefits of cooperation and the feelings of guilt, shame, and anger combined with the pressures toward consistency within discussion that have been selected due to their fitness for facilitating cooperation.

In the picture of human normativity Gibbard develops, the explanatory work is done without recourse to a peculiarly normative faculty that accurately gets hold of a normative feature of the world in some sort of descriptive manner. Normativity involves

---

57 Amelie Oksenberg Rorty criticizes Gibbard for his emphasis on guilt and shame, arguing that these emotions are “neither necessary nor sufficient for the psychology of coordination.” See Amelie Oksenberg Rorty, “The Many Faces of Gibbard’s Wise Choices, Apt Feelings,” *Ethics* 103 (January 1993): 320. And, as Rorty rightly points out, these emotions are severely challenged in terms of their directive force or the selectivity of their objects (323). They are blunt tools. In a similar vein, Thomas Hill asks, “There must be something to motivate the [moral] reactions, but must this be feelings of guilt and anger?” (Thomas E. Hill, “Gibbard on Morality and Sentiment,” *Philosophy and Phenomenological Research* 52, no. 4 (December 1992): 960). Gibbard resists such moves because he is trying to do give a noncognitivist account of moral obligation and duty that salvages something of the “objective pretensions.” Gibbard sees guilt and shame as powerful emotions selected due to their ability to coordinate human interactions and as the most promising candidates for a non-cognitive analysis of guilt and obligation in his admittedly speculative evolutionary psychology. Indeed, Gibbard has recently reaffirmed this view: “What is it to think an act morally wrong, as opposed to just silly or imprudent? Roughly, I propose, it is to think that the act warrants resentment on the part of others and guilt on the part of the person who did it. Specifically moral questions, if this is right, are questions of what moral sentiments to have toward things. At their narrowest, they are questions of what to resent people for doing and what to feel guilty for doing.” (Gibbard, *Reconciling Our Aims*, 16).

nothing more than human expressive activity operating independently of any normative properties in the world that are accurately described in that activity. This has two advantages for Gibbard. First, it avoids a central problem for any descriptivist account, namely, that it will find it impossible to give an account of our motivation to act in accord with our moral beliefs. As seen above, Gibbard thinks this “single loophole remains unpluggable by descriptivistic analysis,” and he labors to plug it with an expressivist analysis.\(^{59}\) Gibbard here is greatly impressed by “moral judgment internalism,”\(^{60}\) and uses it to dismiss descriptivism and motivate much of his own account. Second, and perhaps more importantly, if Gibbard’s account works, it allows him to maintain his Galilean metaphysic. The world harbors no normative facts, and this is good; for normative facts would raise serious difficulties for a Galilean metaphysic. Gibbard thus highlights the clean lines of his metaphysic: “In explaining why we make the normative judgments we do, I found normative facts superfluous.”\(^{61}\) There is no needless multiplication of entities here.

The question that often dogs such clean accounts is how much this metaphysical fastidiousness costs. As Jaegwon Kim rightly notes, “Physicalism cannot be had on the cheap.”\(^{62}\) And the price of Gibbard’s Galilean story, at first blush, would seem to be the elimination of central elements of the phenomenological surface of our normative life, the debunking of the “objective pretensions” of normativity generally, and the moral life


\(^{60}\) “Moral judgment internalism” is the idea that when someone makes a moral judgment they will be motivated to act on it to at least some degree. See Chapter 2, section I.C for more discussion of issues of internalism and externalism. Some of the other points where Gibbard’s moral judgment internalism are close to the surface of his account include, Gibbard, *Wise Choices, Apt Feelings*, 10, 49, 56, and 70.


\(^{62}\) Kim, *Mind in a Physical World*, 120
particularly. After all, according to Gibbard, it turns out that normativity gets nothing right and that it is, in the final analysis, the human activity of accepting this or that norm. It could easily seem that normativity and, with it, morality, leaves us with nothing but various human expressions of this or that norm being accepted. We would then be faced with the question of whether Galilean metaphysical cleanliness is worth the steep price, and if perhaps we should search for a less costly alternative. Gibbard, however, does not believe he has to pay that price in full, and that it would be a significant blow to his account if he had to. We will now look at how Gibbard believes he can have a Galilean metaphysic and something of moral objectivity too.

B. Gibbard on the “Objective Pretensions” of Morality

Gibbard identifies three major pretenses of normative objectivity that he would like to do justice to if he can. But first he makes clear what we cannot salvage out of objectivity. We have a lingering hankering for Platonism and this simply must be outgrown.63 “Norm-expressivism is meant to capture whatever there is to ordinary notions of rationality if Platonism is excluded.”64 Whatever objectivity is about, it is not a matter of getting things right with respect some sort of independent metaphysical reality of a wholly different kind. There are no Platonic “forms” or their ilk floating about. If anyone still seriously holds to such an account, Gibbard wants only to debunk it as being too spooky (that is, smacking of

63 Gibbard’s rhetorical strategy is similar to John McDowell’s in Mind and World and elsewhere where he sets up something he calls “Platonism” and then pushes off against it.

64 Gibbard, Wise Choices, Apt Feelings, 154.
a hankering for the mental to be prior to the material) to fit into the materialism Gibbard takes as given.\footnote{For a recent, sophisticated Platonism of the sort Gibbard is committed to debunking, see Robert Merrihew Adams, \textit{Finite and Infinite Goods: A Framework for Ethics} (Oxford: Oxford University Press, 1999). Gibbard would also be committed to debunking the sort of account this dissertation attempts to motivate.}

Having laid aside what he takes to be the patently unrealistic demands for a “Platonist” type objectivity, Gibbard gets down to the real challenge presented by three ways in which ordinary people do seem to be claiming a sort of objectivity for their normative, and especially moral, judgments. Gibbard hopes to give a plausible materialist account of what he calls the “objective pretensions” of human moral phenomenology. In brief, these are as follows:\footnote{The following points summarize points from Gibbard, \textit{Wise Choices, Apt Feelings}, 155.} (1) people take norms they accept to hold independently of their own acceptance of the norm; in other words, they take it that it would hold even if they were to reject it (e.g., it would still be right to believe that Mao was wrong to kill millions of people even if everyone on earth thought otherwise)—call this the \textit{pretense of independence}; (2) people take their acceptance of norms to owe to requirements of rationality or good reasons, rather than to contingencies of their particular biological or environmental situation—call this the \textit{pretense of non-arbitrary reasons}; and (3) people take their claims that something is rational or morally right to apply not merely to themselves, but other people, possessing an authority that stakes a rightful claim to govern the behavior of others—call this the \textit{pretense of interpersonal authority}. In approaching these three “objective pretensions” of human moral phenomenology, Gibbard wants to avoid a simple debunking project in favor of seeking to preserve whatever can be preserved.
Yet the very phrase—objective pretensions—indicates that Gibbard realizes that he is committed to at least some sort of softening of even these kinds of objectivity that are not implicated in “Platonism.” One might suspect that perhaps this is precisely because these claims to objectivity do in fact reside naturally in a thicker metaphysic than Gibbard can abide. Be that as it may, Gibbard sets about rescuing whatever he can of these central elements of the phenomenological surface of human normative experience, employing a strategy of showing how we naturally find ourselves falling into these ways of thinking and making sense of things, and that this natural proclivity is all that we need. As we think about how to live, this is just what we do; and we find that it works.

Regarding the first pretense, the pretense of independence, Gibbard urges us to see that norm-expressivism nicely explains why we take our normative beliefs to hold independently of us. Basically, if we look at acceptance itself, it consists of thinking that this is the way to think; to accept something is precisely to think that this is what makes sense. As such, the person who accepts a norm cannot but think that this norm applies to others who disagree.67 There is no way to accept a norm and not have it amount to this. In effect, to our “I accept this” there is always an implied “and this is the way to think about this matter.”68

Gibbard’s thought about the first pretense shades into his treatment of the third pretense, the pretense of interpersonal authority, and it is hard to draw a line between them. At the heart of his account of the authority that apparently resides in normativity, Gibbard sees what he calls a “conversational demand”—a sort of Stevensonian “…you do so as well”


68 See the Blackburn quote in Chapter 2, pp. 125-26 above, for an example of this kind of move.
freighted with our endorsement of a norm.\footnote{Gibbard, \textit{Wise Choices, Apt Feelings}, 172-4.} Both of these can be seen in the implied “and this is the way to think about this matter” that rides along with an “I accept $n$,” whatever norm ‘$n$’ happens to be. There is some pressure applied; when I accept some norm and express that acceptance to you, I lean on you. “A speaker,” Gibbard says, “issues conversational demands, and the demands amount to claims to authority.”\footnote{Gibbard, \textit{Wise Choices, Apt Feelings}, 174.} When someone makes a “normative” claim, there is an unavoidable sort of demand, a “you do so as well.” If you accept some other norm and express that, there is a tension that exists between us, something that threatens a potential breakdown of cooperation. You are bucking my “do so as well.” You express a different way of thinking, along with an implied “No, \textit{this} is the way to thing about this matter” with its attending universalizability—a “\textit{you} do so as well.” We would appear to be at loggerheads, but there is a pressure for resolution, for us to work toward a point of cooperation. When things go smoothly, the authority will work itself out in normative discussion that tends toward a resolution that enables cooperation.

Authority, then, resides in the demands one person makes on another in conversation. “Conversational demands amount to demands for influence. To claim authority is to demand influence,” Gibbard states.\footnote{Gibbard, \textit{Wise Choices, Apt Feelings}, 173.} Selection pressures have wired us so that this usually works itself out in discussion, and a “we” is arrived at with an agreed authoritative claim. \textit{We all} accept this norm ‘$n$’. This collective demand has more force—that is to say, authority. One may survive better having joined up with this \textit{we}. As Gibbard
puts it in discussing a case where a particular person accepts a norm that “we” do not, “She, then, considers what she is doing to be rational… The point here is, the rest of us do not.”

The aggregation of demands has added up to greater authority. Authority turns out to be “conversational pressure”—and, as Gibbard notes, this leaves real questions about who is issuing such demands and the context of norms a speaker accepts and attempts to bring to bear in such pressure. These questions will be taken up shortly.

Turn finally to the second pretense, what I am calling the *pretense of non-arbitrary reasons*: when I accept a norm I take myself to do so for principled, rational reasons, rather than because of biological or environmental happenstance. In his attempts to salvage something from the second pretense, Gibbard advances the idea of higher order norms. Gibbard argues that we embrace some norms that are removed from the immediacy of our circumstances, norms we accept that govern *how we accept other* norms in a wide variety of circumstances. We may, for example, accept a higher order norm that demands consistency between the norms we accept, or that a norm be accepted only when we have considered available alternatives. In this case, we have, as it were, a sort of heuristic that we apply across the board when considering which norms to accept. Here there seem to be principles we adopt, and these higher order norms do not seem idiosyncratic, or at least not as obviously so. At the limit, we might find the theory of reflective equilibrium, which Gibbard understands as a system of higher order norms where one considers “vividly all

---


74 Something of a regress threatens, for there is another question of what norms guide the acceptance of the “higher order norms.” At some point in a non-cognitivist analysis there will be a moment of unalloyed acceptance or endorsement unguided by further norms.
relevant facts and philosophical arguments” and arrives at consistent judgments. To be in reflective equilibrium is to be in a state where one has taken all available steps not to be idiosyncratic in one’s judgments—and it seems that nothing else could be asked of a person with respect to this second pretense.

Yet, as Gibbard points out, “we have no guarantee that reflective equilibrium will be the same for everyone.” There is a problem here. It is possible “that both you and I are in reflective equilibrium and we accept incompatible norms. We each recognize the other as logically coherent, and we know we are each in reflective equilibrium.” And then we can go no farther. We are left with an idiosyncrasy of competing efforts not to be idiosyncratic, each consistent from within the ambit of its own situated attempts to be rational. Nor does it help to regard the norms one accepts “as having only a standpoint dependent validity.” Higher order norms do not finally resolve the matter of how there can be something like non-arbitrary reasons. An appeal to “higher order norms” does not salvage much of the pretense of non-arbitrary reasons.

Gibbard believes there is nothing to do but embrace a kind of “parochialism” where we attempt to make our “parish” as wide as possible. We can hope to “take all humanity as our parish,” but when we find that someone simply accepts a system of norms that we do not accept and no amount of discussion results in their acceptance of the norms we accept,

---

75 Gibbard, Wise Choices, Apt Feelings, 170.
76 Gibbard, Wise Choices, Apt Feelings, 170.
78 Gibbard, Wise Choices, Apt Feelings, 170, emphasis in original.
where does that leave us?\textsuperscript{79} All the discussants have done their best to exert conversational pressure and that pressure has been withstood by all concerned. Do we let this fact call our own normative judgments into question? As Gibbard puts it, “We each make our conversational demands. Is there anything left to say?\textsuperscript{80} When we are at loggerheads, “perhaps we must confine our judgments to a smaller community that is indispensible to us or nearly so.”\textsuperscript{81} In other words, we must be content to limit (that is, be “parochial” in) our moral judgments and the conversation partners to whom we will respond as giving us “rational” input to a smaller community relevant to our projects (a “parish”), not concerning ourselves with other communities that are not relevant. Given that different human beings accept different norms and that “content-neutral” avenues of appeal are of limited power,\textsuperscript{82} some sort of parochialism seems unavoidable.

Consider an example that will be developed extensively below: think of an abolitionist in the antebellum South who exerts all the conversational pressure on the slaveholders he can muster, yet the slaveholders see no reason why the enslaved group of \emph{homo sapiens} should be valued in the proposed way. Likewise, the slaveholder makes his conversational demands to no avail, the abolitionist obstinately clinging to norms he accepts. What do they do, each faced with such strange obstinacy? Suppose that one of them investigates the intelligence and rationality and factual information of the other, but (as Gibbard puts it) turns up “no content-neutral qualification he lacks. Would his normative


\textsuperscript{80} Gibbard, *Wise Choices, Apt Feelings*, 192.

\textsuperscript{81} Gibbard, *Wise Choices, Apt Feelings*, 203.

\textsuperscript{82} See pp. 165-69 below.
authority then be vindicated?” Faced with such a prospect, might he not ask: “could we not reasonably dismiss him, simply on the grounds that his normative judgments are monstrous? The alternative,” he might think, “is to leave all our normative judgments at the mercy of a freak normative sensibility.” In short, he might opt to reside in a “parish” where “we” dismiss “them” as fundamentally blind and presently useless as partners in discussion. As Gibbard puts it, “we need, as a last resort, to be able to exclude him. It may well be futile to tell the person himself that he is a bad judge of normative matters. We may, though, need to tell each other. Otherwise we leave our normative convictions unduly hostage to the possibility that a few others—a few quirky people in our midst or many strange people far away—will see things differently.” In the end, what “we” accept can be sheltered from the conversational pressure others bring to bear on “us” by dismissing them as conversation partners: “We pay a price, but the price may be worth paying.” Here we come back to the brass tacks Galilean metaphysic in its evolutionary manifestation: “Sheer coherence places few constraints on who can be left out and why: the constraints with teeth are ones of cost.”

Nor (as will appear below) does Gibbard think that an appeal to higher-order norms will

---

83 Gibbard, *Wise Choices, Apt Feelings*, 196. In this passage and the next, Gibbard is considering the case of “fully and vividly” informed Caligula who lacks no “content-neutral qualification” rather than a slaveholder and an abolitionist each considering the other in the same conditions. All the points still apply to the latter case, I am merely pointing out that they apply *equally to both* given the different systems of norms each accepts. Although “we” accept the same norms the abolitionist accepts (though not, perhaps, for the same religious reasons most abolitionists in the United States accepted them) and so agree with dismissing the slaveholder as a discussion partner, it is important to see that the slaveholder can play the same game and Gibbard has not shown why he has any less right to do so. This point will be developed below in the case of the ideally-coherent slaveowner.


settle matters. The end of the matter for Gibbard has to do with pragmatic issues, a “pragmatism of last resort.” How much will it cost me if I do not cooperate with this person? What difference would it make for the desires I would like to see fulfilled if I no longer treat this person as a conversation partner? What if he has a lot of friends? “Excluding a person has its costs, and including him may too.” In the end, there may be nothing else for it but to write someone off. Gibbard makes clear that one cannot rule out moving openly with such pragmatic measures should one find oneself at conversational loggerheads with a person—and clearly this has happened in at least some discussions about slavery (think of the Civil War). In such cases, this person (or group of people) simply will not accept ‘n’, and I cannot jettison ‘n’ at a reasonable cost, so I work around the person, go through the person, or avoid the person. And then I do damage control. There is, however, only so much damage I can control—thus the “teeth” of this kind of cost. These teeth force me back to working together with some group of people or another, hammering out an authoritative acceptance of norms that “we” endorse and that enable life-giving cooperation. The objective pretenses of normativity are part and parcel of these cooperative efforts.

C. Gibbard’s Problem of Circumscribing Acceptance and the Specter of Moral Reasons Arbitrariness

In what follows it will be argued that Gibbard fails to capture the objectivity so obviously central to human moral phenomenology, and that his account founders on problems embedded in his understanding of moral reasons. With rationality and reasons


having been made into a matter of endorsement, what will limit endorsement? Rationality cannot constrain acceptance because acceptance constitutes rationality, and Gibbard analyzes normative reasons in terms of his notion of rationality as acceptance; so there is no help from that corner. What will constrain acceptance? Gibbard’s answer: A psychology grounded in evolutionary selection pressures that have selected for the emotions and thought processes that foster cooperation. Gibbard includes “higher order norms” within that evolutionary psychology, but he insists (rightly or wrongly) that higher order norms of consistency, full and vivid information, and other putatively “content-neutral” constraints cannot shoulder much of the burden by themselves. Finally they need—and were themselves selected because of—evolutionary pressures that push us to “widen the parish” to include others. Gibbard’s solution, however, can be seen to be inadequate to the task at hand—or so it will be argued in what follows.

To understand the deep difficulties in Gibbard’s story about normativity, it will be helpful to begin with his account of reasons. To come right to the point, reasons turn out to be expressions of acceptance of some norm or other. “When a person calls something—call it R—a reason for doing X, he expresses his acceptance of norms that say to treat R as weighing in favor of doing X.” Reasons, for Gibbard, do not stand apart from acceptance; they express the acceptance. When someone offers a reason “she thereby expresses her

---

90 Gibbard elsewhere put this as a matter of our “proclivities” (Allan Gibbard, “Morality as Consistency in Living: Korsgaard’s Kantian Lecture,” *Ethics* 110 (October 1999): 146). Gibbard is taking issue with Korsgaard’s proceduralism and advancing a notion of a substantivalist stopping point for “why” questions. Specifically, Gibbard sees our evolutionarily conditioned psychologies and the “proclivities” they give us as playing an absolutely central role in ethics.

acceptance of a system of norms.” If this account is right, then reasons do not constrain acceptance; reasons just are that acceptance come to expression and thus themselves stand in need of constraint if our normative life is not to collapse into a naked competition between instances of acceptance of one or another system of norms. Something other than reasons must constrain acceptance of norms. “Discussing reasons, then, has left us where we started on objectivity,” Gibbard observes. And so the same question presses still more forcefully: What becomes of normative objectivity when all the reasons we offer merely express an attitude of acceptance of a system of norms?

Staying with this issue of reasons for a moment, another way to see the problem is to consider what goes on when one person dissents from another person’s reasons for doing something. “When a person does something we think ill considered, we distinguish ‘his reasons’ for doing it from ‘a reason’ for doing it.” For Gibbard, this distinction can only stem from each of the parties involved happening to accept different systems of norms. Rationality and reasons have no traction here, no ability to adjudicate the disagreement; they simply express the disagreement itself. “We” look at his “reasons” and find ill-motivated considerations that don’t add up to what we would accept as a reason. But he squints right back at us and returns the compliment. Our “reasons” are nothing of the sort from within the system of norms he accepts. The parties here are embedded in two different contexts of acceptance, and nothing specifically normative remains to be said between them. That is to say, offering reasons (which now means: “We accept this; you do so as well”) amounts to

92 Gibbard, Wise Choices, Apt Feelings, 163.

93 Gibbard, Wise Choices, Apt Feelings, 164.

94 Gibbard, Wise Choices, Apt Feelings, 162.
browbeating (that is, trying to persuade one’s listeners by louder and more insistent appeals to a normative system they don’t accept) unless someone is first cajoled or joshed or wooed or beaten into the other system of norms—or perhaps the offering of “reasons” just is such cajoling and pressuring.\(^{95}\) Offering our best “reasons” in situations of fundamental disagreement would then amount to a restatement of the disagreement, and the other party to the disagreement will not see such reasons as good reasons, as reasons that \textit{ought} to motivate us. And if they do respond to those “reasons,” they will not be responding to them \textit{as reasons}, but on an altogether different level.

Gibbard drives home the limitations of reasons by introducing an example that he develops at length: “Suppose a woman were to starve herself to death for the sake of a trim figure. She acts irrationally, I would say.”\(^{96}\) Gibbard quickly shifts to the first-person plural in speaking of a preference for life over anorexia, noting that “we maintain that this preference is not just a matter of taste…. Our so thinking consists in \textit{accepting} a norm that says without qualification, ‘Do not starve yourself to death for the sake of a trim figure’.”\(^{97}\) Consistently with his non-cognitivism, Gibbard makes \textit{acceptance} of a system of norms central to the issue between the anorexic and non-anorexic. Gibbard then adds a crucial twist that indicates just how deep his commitment to non-cognitivism runs: “Consider an idealized anorexic—one who is acting not against her \textit{reflective} preferences, but on them.”\(^{98}\)

Having reflected on her preferences, the “ideally coherent anorexic \textit{accepts} norms that

\(^{95}\) The language here is purposefully reminiscent of Richard Rorty; it is not clear to me that Gibbard offers much more of substance than Rorty does in these matters. Gibbard’s thoughts on “browbeating” are in \textit{Wise Choices, Apt Feelings}, 191-3 (esp. 193) and 207.


\(^{97}\) Gibbard, \textit{Wise Choices, Apt Feelings}, 165; emphasis added.

\(^{98}\) Gibbard, \textit{Wise Choices, Apt Feelings}, 166; emphasis added.
prescribe death by starvation.”\textsuperscript{99} In Gibbard’s example, the ideally coherent anorexic has reflected on her preferences but has accepted a system of norms that many people do not accept.

Gibbard’s initial point in introducing the example is to argue, in accord with the pretense of \textit{independence}, that we cannot help but regard our moral norms as holding independently of our own acceptance of them. “The norms we accept prohibit starving for a trim figure, regardless of what one prefers or what norms one accepts. For a hypothetical case in which we ourselves are ideally coherent anorexics, these norms prohibit us too from starving. It is in this sense that we think that the prohibition against starving for a trim figure is valid independently of our own acceptance of it.”\textsuperscript{100} Perhaps in something like this performative sense Gibbard is right: our acceptance of the norms means that we accept them as holding universally. (Of course the same holds in the anorexic’s eyes for the norms that she accepts, but let that pass for the moment.) But this sense of independence is very thin; for whatever norms I accept I know that if certain non-moral contingencies had played out differently, I would have accepted a different system of norms and would have viewed \textit{those} norms as holding universally. Having accepted those norms—whatever they are, for whatever reasons I have accepted them—I will view them as holding independently of myself in Gibbard’s attenuated sense.

But the example of the ideally coherent anorexic also shows clearly that the human activity of \textit{acceptance or endorsement} is central for Gibbard’s moral non-cognitivism,\textsuperscript{101} and

\textsuperscript{99} Gibbard, \textit{Wise Choices, Apt Feelings}, 171; emphasis added.

\textsuperscript{100} Gibbard, \textit{Wise Choices, Apt Feelings}, 166.

\textsuperscript{101} See the passages at notes 24, 36, 38, and 39 above.
Gibbard’s further development of this example helps make clear how deep Gibbard’s noncognitivism or expressivism runs. Gibbard ties the example of the “ideally coherent anorexic” into his account of authority as “conversational demands” and into his “pragmatism of last resort.” “Consider again the ideally coherent anorexic. I say that it doesn’t make sense for her to starve herself to death for the sake of a trim figure. She says it does. We each make our conversational demands. Is there anything left to say?”

Gibbard observes that the two parties may each issue epistemological challenges, calling into question the other’s capacity to judge the issue rightly and that it seems “we have arrived at the point simply of pitting my normative authority against hers.” But perhaps an appeal to “higher order norms” might lead out of this impasse: “if conversational demands are to count as legitimate and ideally coherent, they must be backed by an epistemic story. The speaker must accept higher order norms that tell the hearer to accept what he is saying. These higher order norms form his story of why he is in a position to judge, of why his audience should accord him the authority he claims.” Gibbard here sounds as if he is going to head in the direction outlined by full-information descriptivists like Richard Brandt. “We” can appeal to higher order norms and argue that the ideally coherent anorexic hasn’t reflected long enough or has reflected in the wrong way or has violated some sort of higher order norm of ideal rationality or has simply missed some crucial facts, thus failing to act in accord with higher order norms governing her thought. In short, upon further examination, the ideally coherent anorexic turns out not be ideally coherent after all.

102 Gibbard, Wise Choices, Apt Feelings, 192.
103 Gibbard, Wise Choices, Apt Feelings, 192.
If Gibbard were to move in this direction, it might solve the conversational standoff with his “ideally coherent anorexic,” but only at the cost of abandoning his non-cognitivism for a version of descriptivism, the latter being a position that (as was seen above) Gibbard thinks fatally flawed. In the end, however, Gibbard rejects the solution to the ideally coherent anorexic impasse offered by a fundamental appeal to higher order norms by upping the ante on the ideally coherent anorexic.\textsuperscript{105} Take the case of Caligula. “I might think that no one could accept such norms [of gratuitous cruelty] if he fully and vividly understood what was involved in suffering cruelty.”\textsuperscript{106} Since Caligula fails this test rather spectacularly, one might “take it as a sign that he lacked a content-neutral property competent normative judges must have.”\textsuperscript{107} Caligula could thus be dismissed as a competent normative judge since he fails a test of higher order norms. Gibbard thus appears to be following the full-information descriptivist line.

The opening line of the next paragraph, however, refers back to this proposed full-information descriptivist move: “This may seem a reasonable restriction on epistemic stories, but we should nevertheless reject it. Take Caligula again. On confronting him we would at first, doubtless, take his admiring gratuitous cruelty as evidence of some content-

\textsuperscript{105} Indeed, even his account of higher order norms appeals to what an audience will accept (Gibbard, \textit{Wise Choices, Apt Feelings}, 193): “the test is this: could he coherently make his demands, revealing their grounds, and still not browbeat his audience? What makes for browbeating in this test is a question of conversational inhibitions and embarrassments.” As Gibbard recognizes, what will pass this test from case to case will depend on the audience one is addressing. The problem of violating higher order norms of consistency is glossed in terms of “conversational embarrassment.” This point does not need to be developed to make the point that is most crucial in the present context.

\textsuperscript{106} Gibbard, \textit{Wise Choices, Apt Feelings}, 196.

\textsuperscript{107} Gibbard, \textit{Wise Choices, Apt Feelings}, 196.
neutral defect. Suppose, though, we investigate and find no content-neutral qualification he lacks.”  

Gibbard continues on to make his point clear:

Content-neutral properties, after all, logically entail no particular normative judgments. If content-neutral qualification rule out wild judgments, they do so in virtue of contingent limits on the ways people are constructed. These limits need not stem from happenstance, to be sure; we are shaped in systematic ways by natural selection and by life. Still, the limits are in no way guaranteed to protect what most needs protecting. *Caligula, it might turn out, is superb as a normative judge by every content-neutral test.* If his judgments are monstrous, though, we shall not give them weight. If only content-fixed qualifications will rule Caligula out, then we cannot renounce them. We shall dismiss normative judgments if, in various substantive ways, they are egregious.

Gibbard thus urges us to accept that content-neutral higher order norms fail to offer the needed normative constraint. In other words, someone can be consistent, have the facts correct, and adhere to other relevant content-neutral tests, and yet make moral judgments along the lines of Caligula. To rule out such moral judgments, “we” must appeal finally to content-fixed, substantive normative judgments arising from the contingent limits of our evolutionary history: “some of the qualifications a competent normative judge must meet are content-fixed.” Of course we might find that some of those contingent, evolutionary limits do not rule out slavery or being a dictator or playing both sides of an “implicit agreement” if it’s viable to do so. Surely *that* would be an *a posteriori* matter to be looked into by biologists and evolutionary psychologists. In any case, whatever these contingent limits may turn out to be, their content is fixed by non-cognitivist acts of acceptance or endorsement for which we have been selected.

---


Gibbard could, of course, be quite wrong about all this, and no doubt many moral philosophers inclined toward full-information descriptivist theories will be quick to take issue with him about an ideally coherent Caligula (an ideally coherent anorexic was bad enough). That said, Gibbard clearly publishes his commitment to a non-cognitivist moral theory; ideal rationality is clearly not the final court of appeal. That distinction belongs to the non-cognitive activity of acceptance or endorsement.

Having again decisively rejected the full-information family of descriptivist moral theories (and thus what he views as the most hopeful descriptivist account), Gibbard makes his position clear and turns toward his “pragmatism of last resort.” Immediately following his ruminations on Caligula, Gibbard starts talking about pragmatic costs. He does not abandon epistemic coherence totally, but begins talking about it in terms of what it costs one to hold a position others in the community relevant to the success of your projects take to be incoherent. “Dogma,” Gibbard observes, “has a price. Ordinarily it ends discussion. One can be baldly dogmatic toward someone only if one is willing to dispense with him as a discussant. It is this price that gives the requirements of epistemic coherence their teeth.” And this is the final word with the “ideally coherent anorexic.” Again, there are different “conversational demands.” “The facts, suppose, are agreed between us, and so our difference is entirely normative. I know that the higher order norms on which I base my demands are not the ones she accepts. She, of course, can make corresponding demands on me: she can demand that I accept that it does make sense to starve oneself to death for the sake of a trim figure, and she will be

111 It should be quickly pointed out that Gibbard nowhere uses the phrase “ideally coherent Caligula,” and I do not insist on it. Gibbard’s own language in the above passage is sufficiently clear: “Caligula, it might turn out, is superb as a normative judge by every content-neutral test.”

sincere in virtue of the higher order norms that she accepts.” Clearly, acceptance has the final word (as it must, for a non-cognitivist), and if one is willing to pay the pragmatic costs of dismissing another person or group of people as discussion partners or as partners in cooperation or perhaps even in a modus vivendi, then there is nothing more that can be done. We have to be prepared to be “parochial” in our moral judgments, even if we aspire to “take all humanity as our parish.” “For our central normative judgments, we must be prepared to give up the claim that any human being whatsoever would accept them, if only be were ideally placed for normative judgment. Perhaps we must confine our judgments to a smaller community that is indispensable to us or nearly so.” The reasons that community finds coherent will be the ones that matter to us. Clearly the “reasons” one recognizes will depend a lot on which community one finds indispensable.

All of this makes quite clear that reasons have only the most tenuous externality to matters of acceptance. If you do not accept a norm ‘n’ and I do, you have no reason to act in accord with ‘n’ except for the trivial reason that I accept something different. There is a kind of external pressure there, as we have seen in Gibbard’s discussion of authority (to which we will return, below). But while Gibbard perhaps captures a bit of the idea of something external, what goes missing is that what is external to you is a reason. What Gibbard offers is a reason in the same sense that a gun to my head is a reason for me to act. There is a brute force that does in fact have motivational force, but such brute force can be

113 Gibbard, Wise Choices, Apt Feelings, 198; emphasis added. This passage points toward a central element in Gibbard’s later work (explored below, see pp. 241-45) where we find him emphasizing an “idealized Plato” and an “idealized Xanthippe” who agree on all the facts yet disagree about what is “the thing to do.” This and similar examples feature quite prominently in Gibbard, Thinking How to Live.

114 Gibbard, Wise Choice, Apt Feelings, 203.

115 Gibbard, Wise Choices, Apt Feelings, 203; emphasis added.
in the hands of anyone for any cause whatsoever. Perhaps the gun to my head is that of a police officer who would like me to release a three year old girl I am holding hostage. Or, perhaps, it is that of a thug who wants me to give him my wallet. In the former case the force accords with a reason that seems to most people to stand independently of the force applied, while in the latter case the two diverge. The point is that mere force is neutral. Yet it is a central idea of human normativity, and particularly our morality, that force and reasons can diverge. But Gibbard pulls them so close together—with the authoritative prescriptivity of reasons finally collapsing into matters of force or pressure—that reasons become similarly neutral, having no force of their own by which they are able stand over against acceptance and sheer force. Given that reasons are themselves expressions of acceptance, they are not themselves able to constrain acts of acceptance or endorsement. The forces and pressures central to Gibbard’s account of normativity do not respond to normative matters, but shape normativity independently in a way that can only be arbitrary when viewed from within the standpoint of normativity itself. Given his view of reasons as expressions of acceptance or endorsement, reasons do not stand over against or offer support for a person’s choice to accept or endorse a system of norms; they themselves result from such acceptance or endorsement. This makes it difficult to see how Gibbard could avoid moral reasons arbitrariness. Reasons—as expressions of acceptance—are precisely what need to be constrained. They do not perform this function themselves.

If in Gibbard’s story of human normativity, then, reasons do not constrain or guide acceptance or adjudicate competing expressions of acceptance, something else must function in this way if Gibbard hopes to avoid moral reasons arbitrariness. Gibbard needs
something else to ground normativity, something that circumvents the impotence of reasons through a different kind of imposition.

It has already been seen that this constraint arises from the brute facts of how we have happened to evolve and how this plays out in local communities we find indispensable. Gibbard states the end of the matter plainly: “We have all in fact been influenced by pragmatic considerations, and influenced deeply, if my evolutionary speculation is on track. If such influences do not make for good normative judgment, we are hopeless as normative judges.”

It is not clear what Gibbard could have in mind regarding standards of “good” normative judgment that could stand over and above evolutionary pressures, but this much is clear: Whatever limits human acceptance of some system of norms or another, it is not rationality or reasons as such; for these merely express acceptance of the norms in question. The limitations come from brute facts of human psychology, from the motivational sets with which we happen to find ourselves accoutered due to the way evolution played out in interaction with the contingencies of our cultural situation. Nothing stands over against this evolutionary/cultural amalgam. Here we hit bedrock in Gibbard’s Galilean story. The explanation must proceed not in normative terms dealing with normative facts, but in terms that have no normative vision, no “good” outcome in view.

With this evolutionary bedrock in view, we may return to the three “objective pretensions” of human normativity that Gibbard desires to salvage. The pretense of independence and the pretense of interpersonal authority will be examined together. Does Gibbard do justice to the fact that people regularly take their normative judgments to have validity independently of their own act of accepting the norm in question, and what does Gibbard’s

notion of authority as conversational demands preserve of the pretense that normative judgments actually have authority.\textsuperscript{2117} Recall that in dealing with the first pretense, the pretense of independence, Gibbard wants to hold that our acceptance just is an attitude holding that this is the way to think about the matter, that how I am thinking about this now is in fact the way that it makes sense to think about the matter. The very act of acceptance cannot help but reach out to the independence of the judgment from acceptance. There is, we might say in a Habermasian vein, no performative alternative. Likewise, when I express this acceptance, I take a stand that says that this is the way to think about the matter, and that others should do so as well. I make a demand or exert a pressure in conversation that others experience as an assertion of authority: “a speaker… issues conversational demands, and these demands amount to claims to authority.”\textsuperscript{118} The authority of normativity consists in one person demanding that another person accept the system of norms that she accepts, and the independence of norms consists in her sincere acceptance of this system of norms as the way it makes sense to think about the matter, such that others should “do so as well.” Thus the authority and the independence of our norms are intertwined.

All of this takes place in an evolutionary and cultural context. I accept these norms because I need the cooperation of these people who happen to be relevant to my projects. This group is the people with whom I most need to form a “we.” And what if I happen to be a white man living in the Antebellum South or a Hutu living in Rwanda in the early

\textsuperscript{117} Although Gibbard clearly hedges his bets by calling the claims at the heart of normativity objective “pretensions,” thus signaling his willingness to dismiss them outright if need be, he is keen to preserve what he can of them (see notes 3 and 4 above). This is one of the main twists of contemporary non-cognitivists (as with Gibbard and Blackburn) over their predecessors (A.J. Ayer, for example). It will be seen below how difficult this attempt is for the non-cognitivists.

1990’s? The projects of many thousands of moderate Hutus were abruptly cut off due to their failure to accept the system of norms of the more radical Hutu majority whose cooperation they needed. Or again, think of the price paid by the ten Boom family for rescuing Jews from the Nazis.119 The conversational demands of the Nazis were as real as those aimed at the good American boys who fought them. There are all sorts of “we’s” issuing all sorts of conversational demands, and, as Gibbard sees matters, reasons and rationality cannot finally resolve the matter of which authority to accept, since they are themselves precisely expressions of acceptance. They gain purchase only after some system of norms or another has been accepted. Nor (as seen above) does Gibbard believe that an appeal to higher order norms is sufficient to resolve these matters. In the absence of such appeals to a more descriptivist line of thought or to authoritatively prescriptive moral reasons, it is not clear how Gibbard will speak to conflicting conversational demands emerging from different systems of norms sincerely accepted by different groups of people.

What if I need the cooperation of white, antebellum slaveholders for my own projects to succeed? This question will be developed more fully as we progress.

One begins to wonder if Gibbard’s theory leaves matters of authority to the luck of the draw. If the contingencies shake out one way, you will accept one normative system. If, however, things shake out another way, you will accept something else. This creates problems not just for the pretense of authority, but also for the pretense of independence. The person who accepts Gibbard’s story about normativity will think that her acceptance of the particular norms she accepts rests on contingencies of her evolutionary history and her culture. If she starts to think that she is getting something right, a little reflection will

remind her that there are no normative facts that she is getting right and that her views trace back to the contingencies of her situation. In all this we would seem to be much closer to the relativism of someone like Bernard Williams, rather than pursuing a promising route for capturing the objective pretenses of morality.

Moreover, if we grant Gibbard’s picture of authority, don’t we actually have too much of this “authority”? The authorities balance and cancel each other out; in this equipoise we are caught between two “do so as well” injunctions, two demands, two particular instances of acceptance. In each case, their “authority” seems equally grounded, except that one happens to be my endorsed norm and the other happens to be yours. We each unavoidably accept what we accept, but the final authority seems to be nothing other than my own acceptance. Gibbard offers little that can adjudicate competing conversational demands, competing “authorities”; nothing within the ambit of the norms themselves can resolve matters of conflicting authorities. Gibbard seems to face a difficult problem of how resolve such conflicts while remaining within his austere materialist non-cognitivism. If Gibbard stands his ground and does not tacitly appeal to some sort of descriptivist cognitivism, it would seem that only non-normative contingencies can resolve the matter (for example, you happen to have the bigger stick).

By grounding his story of authority in demands and pressures Gibbard is operating with the wrong sort of stuff even to address the issue of authority; my making demands is an interesting sociological fact, but it would seem to have nothing to do with whether my demands are right or should be followed. If my demands carry the day, the sociological facts

120 This is, of course, in contrast to an obvious stance open to a “descriptivist” of some variety, namely, that it is possible to get something independent of all the contingencies correct. Normativity can refer to properties independent of what anyone accepts.
have played out in a particular way, but *nothing at all* has been shown about how things *should* have played out (perhaps Hitler has carried the election). Indeed, for an account where this is all that can be said, one gets a gnawing sense that there *is* no way things *should* play out. Things will just go one way or another, and that’s it. Gibbard fails to transmute force and pressure into authority; he runs a grave risk of confusing the “By what right?” question with demands and pressures. Gibbard’s appeal to conversational demands as the grounds of authority and his attempt to work out the notion of normative independence in terms of sincere acceptance of a *system* of norms do not salvage any noteworthy “objectivity” for human normativity, leaving the *pretense* of independence and the *pretense* of *interpersonal* authority as exactly that: *pretensions*, conceits that are seen through. Nor is it clear what more Gibbard will be able to say without violating the Spartan strictures of his materialist non-cognitivism.

Before moving on to the *pretense of non-arbitrary reasons* and the problem of Gibbard’s pragmatism of last resort, it should be noted that it is already clear that Gibbard fails to avoid the problem of moral reasons arbitrariness. If Gibbard is right, what a person accepts as having authority owes to an act of acceptance that is not finally responding to reasons that address themselves to us as normative beings; instead, our normative reasons turn out to be expressions of acceptance beholden to non-normative forces. The same forces that give rise to our knuckles also give rise to the guilt and shame central to moral normativity, and the same kinds of forces that lead one person to accept the normative system of the group of Taliban radicals relevant to the success of her projects, lead someone else to accept the normative system of Western feminism. What determines the shape of normativity comes from outside the normative realm, and the “reasons” proper to the normative realm turn out to be impotent. The norms we take to be valid independently of our endorsement
of them are explained by our activity of acceptance. The belief that they are independent, while performatively necessary, is simultaneously undercut by the explanation of what is really going on. And authority fares no better. Authority turns out to be a matter of conversational demands that lean on other humans to accept what the speaker accepts (think of the people in white coats of Stanley Milgram’s famous experiment). What is important in a claim to be right is not being right, but the fact of the claim and the force it carries because of the other person’s “do so as well.” In all this, normative justification—including moral justification—turns out to be arbitrary. The real action driving human normativity is always going on someplace else, someplace where normative reasons are simply not in view at all. Gibbard has given us a kind of prescriptivity, but authority goes missing. What we have is a welter of conflicting conversational demands arising from non-normative forces. What one will count as a “moral reason” will be strongly influenced by the group of people whose cooperation one needs if one’s projects are to succeed. All sorts of “reasons” could be prescribed by various groups of people in this manner, but the “by what right” question has dropped out as authority is dispersed into “conversational demands” that any group of people can make. Gibbard gives us normative “reasons” with prescriptivity but no authority, leaving us with no genuine normativity at all. This is moral reasons arbitrariness.

Moral reasons arbitrariness also vitiates Gibbard’s account of the second pretense of objectivity, the pretense of non-arbitrary reasons, as well as the “pragmatism of last resort” that he eventually falls back on. The pretense of non-arbitrary reasons holds that people making normative judgments do not take those judgments to rest in idiosyncratic matters about their biology or culture, but in reasons to which they are responding. At this point, it should
be apparent what kinds of difficulties Gibbard is going to face. The basic problem is that people’s normative judgments are not in fact responding to normative reasons, these being expressions of acceptance of some norm or another; instead, the person’s normative judgment is taken to flow out of forces peculiar to his situation. This is precisely what the second “pretense” wants to deny.

As has been noted, Gibbard considers attempts to circumvent the problem by appealing to higher order norms. But these turn out to be plastic and Gibbard argues that they are quite limited in terms of what they can accomplish. Even the extremely high level norm of reflective equilibrium is compatible with a number of different normative systems. The notion of reflective equilibrium may be thought of “as a system of higher order norms: norms that say to accept whatever norms one would accept if one were in reflective equilibrium. To be in reflective equilibrium is roughly to have considered vividly all relevant facts and philosophical arguments, and to have achieved consistent judgments.”121 But, as Gibbard observes, “notorious puzzles” surround this notion. “Two people may achieve opposing equilibria. If so, then reflective equilibrium theory will tell them to accept opposing norms.”122 Two “ideally coherent” people can accept different systems of norms.123 Moreover, it is not clear why the higher order norms of reflective equilibrium theory should themselves be immune to the basic problem. Western liberal academics like John Rawls may naturally find themselves accepting this higher order norm due to the extreme peculiarities of their situation. Others might accept higher order norms like “Do what will

123 As Gibbard has been shown to hold in the pages above. This will be seen again below as we consider Gibbard’s more recent book, *Thinking How to Live*. 
lead to the most material gain, so long as risks can be managed.” Speaking in purely evolutionary terms, some might argue that this latter norm arises out of a situation not nearly as idiosyncratic as a situation that encourages people to accept a higher order norm of reflective equilibrium.124

Perhaps most damning for Gibbard’s attempts to do justice to the pretense of non-arbitrary reasons is the simple observation of the sheer diversity of what is in fact accepted by people. Certainly there are a lot of idiosyncratic forces at work in the shaping of the norms that people do in fact accept as they ingratiate themselves to wildly varying relevant groups from out of their unique biology. Faced with such wild divergence in endorsed or accepted norms, if endorsement or acceptance is the essence of one’s story about human normativity (as it is for a non-cognitivist), then one will be forced to acknowledge that normativity is in fact highly idiosyncratic in its provenance. Full-information theorists and other moral philosophers can avoid the problem by saying that there is a failure on the part of some people to recognize what is rational or has authority. But if acceptance or endorsement is the final word, this option is not open. The competing systems of norms really are accepted or endorsed by this or that person or group of people.

In the end, Gibbard holds that the higher order norms in question have “only a standpoint dependent validity.”125 But this is exactly what the pretense of non-arbitrary reasons is supposed to avoid. Even given reflective equilibrium theory with its system of higher order norms, the moral reasons we accept owe to a particular standpoint that leads us to accept a particular version of reflective equilibrium. We also cannot help but recognize that

---

124 Although perhaps one could tell a Nietzschean story about reflective equilibrium as being widely accepted because of its status as slave morality, used by the weak and resentful to advance their Will to Power.

if we were in a different situation we would have accepted different norms, and since the final word for Gibbard is a matter of acceptance or endorsement, it’s difficult to see what remains to cut across these competing views to show that one or the other was uniquely right, or rational, or authoritative. Acts of accepting systems of norms truly are relative to one standpoint or another.

Near the end of his attempts to salvage something worthwhile from the *pretense of non-arbitrary reasons*, Gibbard acknowledges that he has opened up a mare’s nest of problems, and he points the reader to a later chapter in his book.\(^{126}\) In this later chapter, a central theme is question of how to negotiate paired problems of “parochialism” (limiting our discussion to “us”) and “relativism” (acknowledging differing norms as appropriate for different people). Gibbard thus speaks of the “Greeks” and the “Scythians,” and supposes that “each group is ideally coherent and ideally informed,” but in fundamental disagreement in the norms each group accepts.\(^{127}\) How should these groups deal with one another? One option is to be “parochial”: “‘We cannot engage in normal discussion with Scythians,’ a Greek might say. ‘On some topics, we can only browbeat them, try to manipulate them, or change the subject.’”\(^{128}\) The other alternative is “relativism”: “‘Our way of life is not yours,’ a Greek may tell Scythian, ‘and the ways of thinking that suit you do not suit us. When we understand each other fully, we shall both see that, for the most part, your ways of thinking are right as they bear on your lives, and ours as they bear on ours.’”\(^{129}\)


There is no need to delve into Gibbard’s machinations with respect to these issues. Suffice it to say that as he thinks through these issues he finds himself driven to what seems to be the bedrock of his theory, a pragmatism that provides “the constraints with teeth.”130 Here we run into matters of cost; norms get cashed out in terms of what they will cost in the evolutionary currency of lost cooperation and increased conflict. It is these pressures that “may lead us to choose a parish, to pick a group to which our judgments shall be parochial.”131 In order for “us” to protect “our” normative agreement from challenge by just any idiosyncratic group of strange people, we may need to dismiss them from our thought. To give an example that will be developed more fully below, “We” slave owners may need to dismiss those weird abolitionists, whose norms are in hock to who knows what odd cultural forces. The slaveholder may recognize that the slave is human and rational, but may value the economic gain and pleasure of controlling another human being. In these situations, groups like the slaveholders and abolitionists “may find [themselves] forced, for want of any better way of settling matters, to a pragmatic standard.”132 Faced with challenges to the normative system they accept in the form of discordant demands from abolitionists that “this is the way to think” about slavery and that the slave owners should “do so as well,” slave owners pressured to accept a different system of norms can opt to become “parochial” and dismiss the abolitionists as conversation partners. When we are at normative loggerheads, matters of cost and benefit come openly to the fore.133 Moreover,


131 Gibbard, Wise Choices, Apt Feelings, 220.


133 Gibbard clearly takes a page out of a rule-utilitarian playbook when he emphasizes that the influence of pragmatic concerns should be indirect. “Direct pragmatic standards could not govern effectively,
even if the slave owners were to respond to the “normative” claims of the abolitionists, they would actually have given in to pressures and demands couched in normative term.

Pragmatic matters—pressures, demands, calculations of viability—dominate as the specifically normative drops from view. Gibbard urges us to realize that such pragmatic concerns have a legitimate place in settling which norms to accept, and that, when combined with an evolutionary story, an “antipragmatic purism” would bring with it “a complete skepticism of the powers of human judgment.”

In Gibbard’s Galilean story pragmatic concerns are the court of final appeals.

Here again Gibbard has led moral philosophy into the coal pit of moral reasons arbitrariness. Moral reasons—a species of the normative reasons that have been reduced to expressions of acceptance of norms—are seen to owe their existence to the idiosyncrasies of one’s evolutionary and cultural situation. The situation you happen to be thrown into will have sway over the norms you accept. This means that our moral reasons are being led about by forces unrelated to normativity except through brute causality. Once again, the justification of moral standards is seen to be arbitrary with respect to the kinds of concerns that dominate the phenomenological surface of our moral life. And Gibbard’s pragmatism only exacerbates the problem. In the place of authoritatively prescriptive moral reasons trading in a normative currency based in our ability to respond to normative reasons as such, we find calculations of costs and benefits in an evolutionary currency based in lost

---

opportunities for cooperation and increased hostility. From within a normative standpoint, these pragmatic considerations appear arbitrary. Consider the dismissal of another group as unworthy partners for discussion. The problem that seems to loom for Gibbard is that everyone gets to play this game. And the more power and clout we have, the more groups we can dismiss; we can do without their cooperation. Such calculation, carried on in pragmatic terms, is wholly irrelevant to matters of normativity as such. One group of people may have the clout or the brute power to dismiss another group of people as conversation partners. But are they right to do so? This question drops out, dissipating into competing expressions of acceptance and demands. Gibbard does not have the resources to ask it from within his theory; he can only ask it as one idiosyncratic individual accepting the norms of the group(s) of people whose cooperation he happens to need. But his biography is no more interesting or relevant than any other person’s biography. Amid the welter of these biographies one is left to wonder if anyone is right—or if it even makes sense to ask the question.

Gibbard might be able to make a way forward with these questions if he abandoned his “norm expressivism” for a full-information descriptivism of some sort, but—as Gibbard sees matters (as seen above)—these theories face their own difficulties. Theistic moral philosophies might also offer a way forward, but Gibbard dismisses these out of hand. For better or for worse, Gibbard is wedded to his expressivist non-cognitivism. In the following subsection some examples will be developed along the lines of the “contented criminal” and the “viable dictator” that attempt to show that Gibbard’s marriage to expressivism is “for worse.” In places where Gibbard should have authoritatively prescriptive moral reasons to offer, he has only prescriptions backed by demands and pressures—and these are
susceptible of being employed by disputants on all sides with equal “right,” so far as we can tell from within the ambit of Gibbard’s own moral theory.

D. Anorexics and Criminals, Dictators and Ruminators

In Chapter 2 two nefarious characters were introduced: the contented criminal and the viable dictator. Gilbert Harman’s contented criminal feels no compunctions against robbery and murder and the like, and—like the ideally coherent anorexic—apparently not due to a lack of rationality or the inability to follow out certain lines of reasoning. His contentedness in a life of crime owes not to any lack of capacity for rationality or failure to exercise that capacity. Nor does he lack knowledge of the facts. He realizes the suffering he causes, but is unmoved. In short, an appeal to “content neutral” norms fails to resolve matters. A similar character, the viable dictator, makes an appearance in Richard Brandt’s influential book, *A Theory of the Good and the Right*. The viable dictator is one who has calculated that being a dictator is a viable option. She believes that by bamboozling, managing, buying off, wholesale murder, and a variety of other tactics, she will actually be able to install herself as a dictator. The viable dictator calculates that all the pragmatic costs—the ones with teeth, as Gibbard puts it—can be managed. In what follows, it will be asked what resources Gibbard’s norm expressivism has for discouraging the formation of contented criminals and viable dictators among those who accept a Galilean metaphysic and believe that human normativity fits into that metaphysic in the way Gibbard describes.\footnote{Brandt, *A Theory of the Good and the Right*, 208, 214-23.}

\footnote{The question being pursued here is not the ability of a metaphysic and attendant ethical view to constrain those who do not accept it (for example, the ability of Christianity to inhibit the formation of viable...}
It is important to remember that the contented criminal and the viable dictator are real-life people and not merely instances of philosopher’s armchair counterexampleing.

There are in fact dictators who have correctly calculated that being a dictator was viable, and who apparently found that their preferences were satisfied in being a dictator. Moreover, as Harman observes (see Chapter 2), there are criminals who do not appear to lack normal human capacities for rationality and who insist that they are exercising them rightly in plotting their course of action (though none of us is ideally rational or coherent, and it is hard to say who has gone wrong and where the fault may lie). Moreover, in both these cases, one need not consider only the extremes. As was mentioned in Chapter 2, there are viable petty dictators of the office or a domestic situation and there are contented tax evaders. There seems to be a continuum on both these scales rather than a simple on/off binary. And probably most, if not all, of us fall somewhere along the scale: For how many people are calculations of viability not a central factor in determining speed on an Interstate?

dictators and contented criminals among those who accept a view like Gibbard’s), but rather the ability of each such view to constrain its own adherents when taken seriously by them.

137 That there have been dictators who—for all we can tell—were satisfying many of their desires is clear from history. From Nero to Mao and many in between (and before) many have successfully calculated that being a dictator or a tyrant was viable, and believed that following this path would best satisfy their desires. Such people might have desires that we would not care to satisfy, or at least not in the same ways (though, as mentioned in the second chapter, there are also petty dictators of the office or house). Perhaps one would want to defend a more objectivist theory of value. That, of course, is not an easy task, especially for the materialist, for whom the Galilean world cannot speak to matters of value (or, if it does, speaks in a rather repugnant way): “That something is in our nature does not endorse it, for horrifying things too are parts of our nature,” (Gibbard, Wise Choices, Apt Feelings, 288). Though it is not an easy task, I think the right sort of theist would be able to make a good case for it, making full use of the resources of theism. See Carson, Value and the Good Life, 215-267 for a well-worked out example. As a materialist, however, Gibbard cannot pursue a specifically theistic tack like Carson’s (though it should be noted that Carson suggests that a materialist might be able to employ a modified version of his theory). To enter into these questions fully would move well outside the scope of this dissertation. Gibbard, in his central work on norm-expressivism, does not work out a theory of value (the word “value” does not even appear in the index), though it seems likely that he would be inclined to accept a subjectivist theory. And, indeed, when Carson and Peter Railton push him on this issue, Gibbard replies: “briefly, good means desirable,” (Gibbard, “Reply to Blackburn, Carson, Hill, and Railton,” 980). As I see it, Gibbard has little choice in this matter, given his other commitments—but, again, to argue this would take the dissertation quite far afield.
How many of us *never* steamroll someone else just to get our own way? Perhaps there is something of the impulse of the dictator in each of us.

Gibbard introduces his readers to two close cousins of the contented criminal and the viable dictator, namely the ideally coherent anorexic and the private ruminator. The example of the “ideally coherent anorexic” has been developed above, and here I only remind us of the example and its importance. The ideally coherent anorexic is the person who accepts a system of norms that prescribe starving to death for the sake of a trim figure, and whose higher order norms prescribe accepting these norms.138 Given her acceptance of a particular system of norms, she is both sincere and impeccably coherent in drawing out the conclusions for how she lives her life. Having accepted the normative system, everything else falls out naturally from there, with no gaps in her working out that system into a coherent whole. Like the contented criminal, norms are accepted that many other people do not accept, but, in both cases, there is nothing about the facts of the matter that allow the rest of us to launch an immanent critique of her way of living. Elenchus cannot help. “The facts,” Gibbard supposes, “are agreed between us, and so our difference is entirely normative.”139 That is to say, it comes down to a matter of different instances of a fundamental act of accepting one system of norms rather than another.

Gibbard’s reflections on the ideally coherent anorexic reveal a great deal about norm expressivism, for it is at this point that we encounter a fundamental disagreement and see the difficulties it poses for norm expressivism. Having accepted the norms she accepts, why should the ideally coherent anorexic care that “we” disagree with her? Since she is coherent,

---


we cannot offer a rational critique. Since reasons are analyzed in terms of acceptance, we can offer her no reasons that she will recognize as such. And the person who accepts Gibbard’s theory sees that any such insistent “reason-giving” can only amount to browbeating or bringing pressures to bear that are clearly adventitious to normative discussion as such. I make conversational demands on her, but these do not address normative reasons to her as an agent, but lean on her as a being in which other forces are in play in determining how the normative surface takes shape.\textsuperscript{140} Of course she can respond in kind, setting her act of acceptance against mine and pitting her conversational demands against mine.\textsuperscript{141} When I claim “that her norms are crazy, so that my own are, in a fundamental way, more to be trusted than hers,” she can respond in kind.\textsuperscript{142} When I challenge her and invoke an “epistemic story” that justifies my claim that her norms are crazy and gives reasons why she should accept my authority as a superior judge of the matter, she can once again turn the tables on me. And, what is important to see, in a situation where rationality is acceptance of norms in the first place, any such epistemic story can only repeat the controversy at a new level.\textsuperscript{143} There is nothing for either of us to get right; we are always pitting one instance of accepting norms against another such instance. Our “reasons” always presuppose the act of acceptance. So, too, for the contented criminal.

\textsuperscript{140} In Kant’s terminology, the conversational demands are “pathological.” See Immanuel Kant, \textit{The Groundwork for the Metaphysics of Morals}, trans. and ed. Mary Gregor (1785; Cambridge: Cambridge University Press, 1998), 13, 25 (\textit{Ak.} 4:399, 4:414).

\textsuperscript{141} As Gibbard puts it, “We each make our conversational demands. Is there anything left to say?” (Gibbard, \textit{Wise Choice, Apt Feelings}, 192).

\textsuperscript{142} Gibbard, \textit{Wise Choice, Apt Feelings}, 175.

\textsuperscript{143} Recall Sturgeon’s observation, quoted at note 1 above.
The problem here is that each of us claims to “see” something the other simply misses, appealing to “a kind of fundamental authority, an authority that does not stem from any common acceptance of more fundamental norms.”\footnote{Gibbard, Wise Choice, Apt Feelings, 175.} The norms, the communities we want to enlist for our projects, the fundamental acts of acceptance—all these fail to give common ground for deciding the matter. Gibbard is left to wonder, then, about the plausibility of your accepting a different set of norms on the basis of my fundamental authority where there is no epistemic story to be told to establish my qualifications as an authority that is not already implicated in the norms in question. “Should anyone… ever accord someone else an authority that is fundamental?”\footnote{Gibbard wonders whether what he calls “fundamental authority” (as opposed to “contextual authority,” where someone else is taken to reason better from premises I also accept) is even coherent: “Perhaps all fundamental authority is bogus,” (177). Basically, Gibbard finally concludes that it is a performative necessity that we accord fundamental authority to others in order to trust our own judgments (176-81). Perhaps. The real question, however, is which authorities to trust in this blind way now.} If not, why should one person assign weight to the normative judgments of others?

In answering these questions, Gibbard is at his least convincing. The first thing he does is to look for partners in guilt: “The puzzle is not just one for expressivists, to be sure.”\footnote{Gibbard, Wise Choice, Apt Feelings, 176.} In this Gibbard is clearly right; many views do face a fundamental problem here. But there are some other alternatives that do not face the problem in anything like the same way and that may have resources available to reply that Gibbard does not. If for example, a “descriptivist” picture of some sort or another is right (this terminology seems lacking, but it will be followed here, since it is what Gibbard employs), then there is some independent property in principle accessible to both parties that would settle matters if grasped by both
of them. Such a belief might helpfully change the way the discussions were carried forward, and—if there are such properties—there might actually be ways of putting ourselves in touch with them. Furthermore, the descriptivist would be in a very different epistemic situation than someone who accepts Gibbard’s picture of normativity. On Gibbard’s picture, we accept two different normative systems and express matters accordingly. To echo Hume’s language, in Gibbard’s Galilean view the two parties paint the same canvas (the “facts” of the Galilean world) in two different ways, with neither one getting the canvas either right or wrong (what would that mean?); they just go about things differently, and nothing about the canvas can say that one picture is more appropriate to the nature of the canvas than the other. With respect to the canvas itself, it makes no difference if it is Rembrandt or Picasso or kindergarten finger-paints. Those who accept Gibbard’s Galilean expressivism are left to argue that there is nothing outside the competing instances of acceptance to adjudicate the impasse, and that “we” have nothing right that “they” miss. There is only difference, not correct or incorrect—except in the rather uninteresting sense that “we” prefer “our” norms for the not very interesting reason that “we” have accepted them.

Not so for the “descriptivist.” The descriptivist can say that those who disagree have actually missed something. It is not, in the final analysis, a matter of competing acts of acceptance, but of having to do with something that is what it is independently of our accepting it or not; we either get it right or we don’t. There is a right or wrong to the matter and the reasons offered in a discussion of the matter can track or fail to track that reality.

147 This is not to say that grasping this property would be easy, or that the nature of that property or of the source of such properties would not have to be taken into account in attempting to know what is there to be known. It is clear that, if there are any such properties, whatever their ontological ground is does not permit them to be accessed using a method analogous to those employed in the “natural sciences.”
Of course a possible reply to all this is just to set to work with Ockham’s razor or to start in on the standard verificationist-type lines: We only have access to the epistemic realm that is always circling back in on itself and can never be seen to have successfully reached out to some ontological anchor, and so on. Perhaps. But in the present context that is just to change the topic. The point here is that the descriptivist does have a view that seems relevantly different with respect to the issue of addressing fundamental conflicts of normative authority. Authority is not vested in our activity of accepting this or that system of norms, but arises only in getting something independent of us correct. This could be wrong, but it certainly engenders a different way of approaching the matter at hand than Gibbard can offer, and—if true—offers real hope that things can be sorted out at some point.

Descriptivists need not share in Gibbard’s guilt here.

Gibbard’s other approach to conflicts of fundamental authority is to argue that if we accord ourselves fundamental authority, then we ought also to do so for others. “If I trust myself, why not others too? Is there anything special about me as a judge?” Since it is hard to specify anything that is special about me as a moral judge (impossible, really, in Gibbard’s story), the implication is that I should accord weight to others as fundamental authorities. But isn’t this a hopeless mess? Which others? A full-information descriptivist might have a ready reply to this question (their difficulties lie elsewhere), but there is no ready answer for the Galilean norm-expressivist, for whom acceptance or endorsement of a system of norms is fundamental. Should I accept the views of the majority? How widely considered? If not, which criteria do I apply to decide the relevant group? Should this group sometimes be other than the group relevant to the success of my projects? How is

that to be motivated, given Gibbard’s views? When do I follow a prophet or a reformer? Should things be considered in a historical view? Should I consider what this or that past culture thought? How about people of the future? One could go on in this vein, but the point is made. Gibbard does not need to tell us only that we cannot help but accept some fundamental authority, but how to choose among the welter of candidate authorities while not violating the strictures of Galilean norm-expressivism. Far from solving the problem or even pointing us in a hopeful direction, Gibbard only widens the same problem.

In the end, Gibbard only gets out of the jam by entering the pragmatic realm; here we can give some reasons with teeth, where disagreeing becomes costly.\(^\text{149}\) This is where the contented criminal enters back into the picture, for while he creates all the same problems for Gibbard that the “ideally coherent anorexic” does, he also sheds light on the function of the pragmatic realm in norm expressivism. The contented criminal differs from most of us in that he is more willing to manipulate pragmatic concerns directly. He gives us a cost of disagreeing that is not easily ignored, and “we” respond by trying to make his actions very costly indeed. Ethics starts to look very Darwinian, with normative reasons having receded far into the background as we come to the most fundamental matters. While we are fiddling around with categorical imperatives, debating social contracts, and generally earnestly entering into normative discussion, the contented criminal cuts right to the chase. And he points out on an individual level that can be managed fairly well by laws, police, and prisons, what becomes massively problematic on the level of competing groups and our various normative systems.

\(^{149}\) Perhaps, with respect to the scenario presented in note 68 above, we could imagine that there is a big enough “we” who accept the normative system that prescribes not starving for a trim figure that “we” can enact legislation that models must have such and such body fat, thus removing the incentive by leaning a bit harder.
The pragmatic concerns are where the real action is, and we seem to be mistaking
the issues if we take the give and take of moral reasons seriously in its own terms and miss
the fact that the “reasons” express prior acts of acceptance of a system of norms that is not
itself responsive to reasons. At some point we get to the Galilean brass tacks. Reason
giving loses its place as such. The earnest giving of normative reasons could very well seem
to be for the naïve and the panglossian idealists. What counts is applying the pressure of
conversational demands and bringing to bear other forces of the same kind. Normative
reasons as such drop from view (though employing the language may serve a clear hortatory
purpose), while pragmatic matters, with their pressures and calculations of power, come to
the fore. Of course matters of consistency and getting the facts right are involved, but these
activities are thought to be in the service of some prior act of acceptance that owes to non-
moral forces. Here Gibbard is not far from Harman’s talk of “implicit agreements” and
“inner judgments.”

Pragmatic calculations are often calculations of viability: “What can I carry off
here?” The viable dictator then poses a related challenge to norm expressivism, a challenge
related to a character Gibbard calls the “private ruminator.” A brief summary of what
Gibbard is talking about will suffice, since the “private ruminator” is familiar to all of us.
Indeed, “each of us at times becomes a private ruminator, more or less.”\textsuperscript{150} So, what is a
“private ruminator”? The “private ruminator keeps his counsel; he must enter into the
common standpoint or feign it.”\textsuperscript{151} “The private ruminator suspects that when he responds

\textsuperscript{150} Gibbard, \textit{Wise Choices, Apt Feelings}, 301.

\textsuperscript{151} Gibbard, \textit{Wise Choices, Apt Feelings}, 301.
to moral demands, he is being taken in.”\textsuperscript{152} Indeed, it would seem that the norm expressivist who takes her own views seriously would be tempted to think precisely this. After all, when I accept someone else’s normative reason I am not responding to something \textit{correct} in this person’s thought about a matter (that reason merely expressing what she accepts), but rather I am responding to conversational demands made by someone whose cooperation I would like to secure or whose well-being for some reason matters to the group relevant to the success of my projects. In Gibbard’s view, when I accept the moral reason offered by another person, something else is really going on—and one could easily start to see this as being taken in. “In his suspicions, [the private ruminator] thinks he responds too much to the interests of others—more than makes sense.”\textsuperscript{153} But, having once seen through normative injunctions to the guilt and the anger behind them, the private ruminator is freed up to ask a question: “why not act for his own advantage apart from all consideration of guilt and anger?”\textsuperscript{154} The question could start to creep in: Would I not benefit more if I could play both sides? I could easily start to think of things in terms of viability, seeing that this seems to address things at the most basic level.\textsuperscript{155} And this is really the central issue here: my relationship to moral reasons changes. Fundamentally, the private ruminator has

\textsuperscript{152} Gibbard, \textit{Wise Choice, Apt Feelings}, 302.

\textsuperscript{153} Gibbard, \textit{Wise Choices, Apt Feelings}, 303.

\textsuperscript{154} Gibbard, \textit{Wise Choices, Apt Feelings}, 301.

\textsuperscript{155} It was seen above that Gibbard does not advocate a direct pragmatism of calculating interests, but rather an indirect pragmatism of working within the broad outlines of a human psychology that has been selected to secure cooperation. It was also pointed out the similarities between this issue of direct and indirect pragmatism and the questions of act utilitarianism versus rule utilitarianism where Gibbard cut his philosophical teeth. \textit{Mutatis mutandis}, many of the arguments employed on either side of the utilitarianism question could be applied to the corresponding sides of the pragmatism question.
taken a step back from the give and take of normative reasons; even where she moves parallel to the reasons given, she is not acting from out of those reasons.

In this the viable dictator is like the private ruminator; but there is a difference. Calculations of viability extend not only to questions of what can safely be concealed from others so that they will continue to play the game the dictator needs them to play, but also to questions of where the dictator can safely move openly. Whose favor does she still need? Here the viable dictator will maintain some semblance of mutuality. But there are other questions the dictator will ask as she calculates what is viable: “Whose cooperation do I no longer need? Which groups can safely be ignored at this point? Which roadblocks to my interests can be eliminated?” Here the dictator will move unilaterally, her self-interest out in the open. A “private ruminator” with power need not stay private in all situations. If the viable dictator reasonably calculates that she no longer requires the cooperation of certain groups whose demands are costly, why shouldn’t she dispense with the pretenses?

Entrenched in a Galilean metaphysic with acceptance as the central normative reality (and this driven primarily by evolutionary forces), Gibbard doesn’t have much to offer by way of reply. If the dictator is a norm expressivist, he must hope that there are enough of “us” who disagree with her to stay her hand. There are no independent, authoritative prescriptive moral reasons for her to be aware of or to be pointed out to her.

**E. Slaveholders and Abolitionists: A Test Case for Gibbard’s Moral Philosophy**

Before moving on to consider Gibbard’s more recent work in moral philosophy, it will be helpful to consider a test case that was mentioned above. If in a clear moral case Gibbard’s norm-expressivism cannot render a moral judgment based in moral reasons that
are in some substantive way independent, non-arbitrary, and authoritative, his moral philosophy will not successfully salvage a significant objectivity from the “objective pretenses” of normativity, succumbing to moral reasons arbitrariness. Gibbard’s norm-expressivism would be seen to be a kind of moral non-objectivism. This subsection develops such a case at length, striving to apply the various elements of Gibbard’s moral philosophy fairly and fully (which cannot avoid a bit of review of points developed more fully above) to a clear moral case.

One clear moral case is that of slavery. Slavery is morally wrong, and for our example of a clear moral case we will consider the kind of race-based, chattel slavery practiced in the United States less than 150 years ago. If Gibbard’s norm-expressivism cannot give a clear objectivism in this case, with the result that normative judgments that such slavery is wrong fail to be independent, non-arbitrary, and authoritative, then norm-expressivism is seriously flawed. Gibbard insists that if his “picture of normative inquiry is roughly correct, the upshot for morals won’t be wildly unfamiliar.” Failure to underwrite the needed moral objectivity in the present case, however, would render the upshot of Gibbard’s picture for morals wildly unfamiliar. As Gibbard’s moral philosophy is applied to this case the feeling of unfamiliarity does slowly begin to gnaw on one.

At the outset of considering how norm-expressivism would handle the case of slavery, it is crucial to remind ourselves that Gibbard takes the austere “Galilean” materialism he believes in very seriously, and that his notion of “rationality” is a term of art tailored to fit this view of the reality. “Our normative life is a part of nature,” Gibbard declares, and he is clear about what “nature” is: “The Galilean core is our story of nature; it

is our rough story of how, ultimately, a wide range of things are matters of fundamental physics. The story includes chemistry, molecular biology, and the theory of natural selection.\textsuperscript{157} “Now, as we begin to expand the Galilean core to include a view of human life, we must develop concepts that are non-normative.”\textsuperscript{158} Accordingly, “apparent normative facts will come out, strictly, as no real facts at all; instead there will be facts of what we are doing when we make normative judgments.”\textsuperscript{159} For a non-cognitivist like himself, Gibbard observes, “there are no peculiarly normative facts,” but rather an “evolutionary account” in which, if it is on the right track, “our normative capacities can be explained without supposing that there is a special kind of normative fact to which they typically respond.”\textsuperscript{160} In accepting a norm, one does not recognize some sort of peculiar fact—say, the wrongness of slavery—but accepts norms in particular ways explainable in terms amenable to “Galilean” analysis. Thus, referring to our acceptance of norms that render it “wrong to torture people for fun,” “we can explain how we come to think this without citing, in our explanation, the wrongness of torturing people for fun.”\textsuperscript{161} In keeping with this materialist picture, then, “the state of accepting a norm, in short, is identified by its place in a syndrome of tendencies toward action and avowal.”\textsuperscript{162} Gibbard is out of all sympathy with a picture of human moral life that involves recognizing moral facts as

\textsuperscript{157} Gibbard, Wise Choices, Apt Feelings, 24, 123.
\textsuperscript{158} Gibbard, Wise Choices, Apt Feelings, 123.
\textsuperscript{159} Gibbard, Wise Choices, Apt Feelings, 23.
\textsuperscript{160} Gibbard, Wise Choices, Apt Feelings, 107.
\textsuperscript{161} Gibbard, Wise Choices, Apt Feelings, 124.
\textsuperscript{162} Gibbard, Wise Choices, Apt Feelings, 75.
normative facts—facts that themselves provide authoritatively prescriptive moral reasons. To use Mackie’s language, such moral reasons would be “queer” in a materialist world. Thus, whatever may be said about the wrongness of slavery, one may not—within Gibbard’s Galilean world—hold that its wrongness consists in getting a normative fact wrong. There are no such facts.

Now if there were a normative fact about the wrongness of slavery, this would certainly provide much more robust resources to account for the independence and the non-arbitrariness of normative judgments that slavery is wrong. Perhaps it could also provide resources accounting for the prescriptive authority of such judgments, say, if there were a Being who, because of His perfectly whole personhood (that is, His integral, holy love) Himself necessarily formed the standard for all personal beings who share in that nature. Indeed, “if there were a God who possessed all of these characteristics, then God’s preferences should be authoritative for us.” If a being who from all eternity exemplified all the perfections of personhood in a perichoretic dance of loving relationship created other personal beings to imitate such relationship, then slavery would be independently, non-arbitrarily, and authoritatively wrong, both as a violation of the personal nature of both the slave and the slaveholder, and as contrary to both the preferences and the prescriptive commands of One who eternally exemplifies personhood worked out in loving relationship. Slavery would be objectively wrong in such a case—in the strongest terms imaginable. But of course a “Galilean” materialist, like Gibbard, will want no part of any of

163 Paul K. Moser’s latest work develops ideas in this vein at length. I think there are new resources in these thoughts for solving the infamous Euthyphro problem, but—again—this would move us far afield from this dissertation.

164 Carson, *Value and the Good Life*, 251.
that. If Gibbard’s metaphysical commitments are right, there are no normative facts, and the objectivity of normative judgments that slavery is wrong cannot be anchored by these will-o-the-wisps.

A materialist descriptivism would be more amenable to Gibbard, and he flirts with what he regards as the most attractive version of such theories, the full-information/ideal-rationality account advanced by his first teacher, Richard Brandt. Rationality obviously plays a central role in such theories, and it is here that Gibbard is at his most slippery, his most Byzantine, and—it should be said—his most tendentious. It is crucial to keep in mind that Gibbard’s notion of rationality is a term of art forged in the fires of his “Galilean” metaphysic. Gibbard’s stubborn insistence on using the term “rational”—with all its power as an evocative trope—while emptying it of its normal meaning, makes exegeting Gibbard’s moral philosophy extraordinarily difficult. One is constantly lulled (and perhaps lured) into forgetting that norm-expressivism is a non-cognitivist theory. Thus, it is all the more crucial to pin down Gibbard’s slippery language before considering its bearing on the case of slavery.

The baseline for treating Gibbard’s non-cognitivist analysis of “rationality” is to remind oneself continually that the analysis is non-cognitivist. As Gibbard puts it, “the analysis is non-cognitivistic in the narrow sense that, according to it, to call a thing rational is not to state a matter of fact, either truly or falsely.” And so we find that, “to call a thing rational is to endorse it in some way,” and that Gibbard “shall be using the learned term ‘rational’ in this broad way. It carries a kind of direct and flavorless endorsement.”

---

165 Gibbard, Wise Choices, Apt Feelings, 8.

166 Gibbard, Wise Choices, Apt Feelings, 6, 7.
again, “to call something rational is to express one’s acceptance of norms that permit it.”\textsuperscript{167} As Gibbard sums it up, “My real claim is not for the word ‘rational’, but for a meaning I want to exploit.”\textsuperscript{168} It is of first importance to keep in mind then that when Gibbard says “rational,” we are to read “direct and flavorless endorsement,” where that is understood in a non-cognitivist way that Gibbard views as uniquely compatible with the austerities of Galilean materialism.

Within the speculative evolutionary psychology that features so prominently in \textit{Wise Choices, Apt Feelings}, “rationality” shows up in ways that make it compatible with a Galilean materialism that—as Gibbard sees matters—has little or no room for a more robust notion of rationality that accords better with what is typically meant by that term. “According to any expressivistic analysis, to call something rational is not, in the strict sense, to attribute a property to it. It is to do something else: to express a state of mind. It is, I am proposing, to express one’s acceptance of norms that permit the thing in question.”\textsuperscript{169} So, to call something irrational is not to say that some property is missing, as it would be on the descriptivistic analyses he explicitly rejects.\textsuperscript{170} “My own analysis, though, is not directly a hypothesis about what it \textit{is} for something to be rational at all. It is a hypothesis about what


\textsuperscript{168} Gibbard, \textit{Wise Choices, Apt Feelings}, 49.


\textsuperscript{170} As Gibbard notes, immediately after the preceding passage; “This may seem perverse. Surely a descriptivistic analysis would be better; that is to say, if a person calls something rational, it would be best to hear him as describing it, as ascribing a property to it. Then we could assess what he has to say and have clear grounds for judging it true or false—whereas on an expressivistic analysis, we can only react,” (9-10).
it is to *think* or *believe* something rational, to *regard* it as rational, to *consider* it as rational.”

Gibbard’s analysis of “rational” eschews the recognition of some sort of property that certain things have and others do not, since—as Gibbard sees things—such properties would not fit well in a “Galilean” world. Instead, Gibbard means to “psychologize” the question and analyze “rational” as a *psychological state* on the part of the being that is accepting something as “rational.” In accord with the “Galilean” metaphysic, the question of what it is for something to be rational drops out, and we are instead offered an analysis of the *psychology* of advanced primates who tend to *regard* or *consider* something to be “rational”—that is, to accept it or avow it consistently with one another.

Given this non-cognitivist analysis of “rationality,” Gibbard has further narrowed how the normative judgment that slavery is wrong can be understood as independent, non-arbitrary, and authoritative—that is, as objective in some meaningful sense. For starters, he clearly will not allow the normative judgment to be uniquely rational in virtue of having a property that other normative judgments (like “slavery is right”) lack. There is no such property, according to Gibbard. Instead, the normative judgment that “slavery is wrong,” taken as a statement that is “rational,” is about the psychology of humans and what it is for them (or some of them) to “regard” such a statement as “rational” or to “consider” it to be such. The matter is “psychologized” in accordance with Gibbard’s Galilean materialism and made compatible with the non-cognitivism Gibbard takes to fit best with that commitment. Thus, in saying that “slavery is wrong” is “rational,” Gibbard expresses a “direct and

---


173 For Gibbard’s account of “acceptance” as consistent avowal, see Gibbard, *Wise Choices, Apt Feelings*, 71-75.
flavorless endorsement” of a system of norms that prescribes that slavery is wrong, where this is understood in a non-cognitivist manner. To state that “slavery is wrong” is “rational” is “not to state a matter of fact,” but to make an observation about what the psychology of (at least some) *homo sapiens* leads them to endorse.

Obviously, Gibbard’s non-cognitivist analysis of rationality would make it exceedingly difficult to hold a full-information/ideal-rationality kind of descriptivism, and, as seen above, he explicitly rejects such theories, devoting a portion of his first chapter to a critique of Brandt’s theory and taking periodic digs at such theories throughout the rest of *Wise Choices, Apt Feelings*. Matters of full-information and consistency cannot be used to dig an “ideally coherent anorexic” out of the system of norms she accepts, nor to dispense with a Caligula who, “it might turn out, is superb as a normative judge by every content-neutral test.”174 As Gibbard sees matters, “sheer coherence places few constraints on who can be left out [of ‘our’ community of normative judgment] and why; the constraints with teeth are the ones of cost.”175 Gibbard thus turns aside from full-information/ideal-rationality types of descriptivism, and in so doing again narrows the possibilities for salvaging some sort of normative objectivism from within his “Galilean” sensibilities. As Gibbard sees matters, an abolitionist’s belief that “slavery is wrong” does not track some property that makes the belief rational; no matter of fact is being stated in calling this belief rational. Rather, one is directly and flavorlessly endorsing or accepting a system of norms that prescribes this.

Moreover, descriptions of content-neutral matters of consistency and full-information do

---


not necessarily favor an abolitionist who accepts a system of norms where “slavery is wrong” is prescribed. Such matters offer few constraints on normative judgments.

To this point, then, Gibbard’s norm-expressivism has offered little reason to think that the normative judgment that “slavery is wrong” has any meaningful sort of objectivity, and Gibbard has shut down two possible tacks that might have offered a way forward. The normative judgment that “slavery is wrong” does not gain independence by being tied to normative facts or facts of rationality, since there aren’t such things. Nor does it gain authority from uniquely tracking such facts, nor from emerging from normative judges who are more fully informed or more consistent. Specifically prescriptive content must be introduced, and this comes only from non-cognitive acts of endorsement or acceptance that Gibbard obstinately insists on calling “rational.” Thus “rationality” is a matter of what various groups of human beings endorse or accept in accordance with their evolved human psychology and their cultural setting. And rather than vouchsafing non-arbitrariness for the judgment that “slavery is wrong,” Gibbard seems to have exacerbated the problem. Certainly he has cut himself off from two possible ways forward (though perhaps there is more to be said for one or the other of them than Gibbard allows). What is more, his norm expressivism seems mired in a morass of at least possibly competing moral psychologies, with little, if anything, that might be able to adjudicate the differences.

What does Gibbard offer in the face of these difficulties that might show that the normative judgment that “slavery is wrong” is non-arbitrary, independent, and authoritative? Gibbard’s treatment of the “objective pretenses” of normative judgment has already been considered at length above, but it will be helpful to examine his efforts in the context of a specific moral problem where there should be an objectivity to our normative judgments.
The first thing to note is that Gibbard is clear that even though matters of consistency do not render one act of acceptance of a system of norms uniquely rational as compared to other acts of acceptance, he clearly retains a role for consistency in moral judgments. Consistency plays a key role in Gibbard’s analysis of “acceptance” in terms of avowal: “as a first approximation, we might say that to accept a norm is to be prepared to avow it in normative discussion.”\(^{176}\) Of course sometimes people avow things that they don’t accept, so this won’t do. But in entering into normative discussion, a person subjects himself to pressures to be consistent, because only as people comply with these mutual demands is there “any hope of reaching consensus in normative discussion. A person who refuses these demands must therefore be a poor candidate for cooperation of any kind—and in human life, cooperation is vital. It is fitness enhancing, then, to stand ready to engage in normative discussion, and so to accept the demands for consistency that are part of the package.”\(^{177}\) Here again we see that pragmatic concerns provide the real “teeth” in matters of consistency.\(^{178}\) Life will not go well for one who will not enter into the give and take of normative discussion by trying to affirm and deny the same things consistently, and by being willing to be persuaded by those around him, even as he attempts to persuade them. “Given some cues, an intrinsic concern with normative matters has been evolution’s cheapest, most reliable way to promote the reproduction of a person’s genes.”\(^{179}\) Pressures to be consistent are an unavoidable part of this.


\(^{178}\) See notes 86-88 above.

Moreover, one cannot simply dismiss others as partners in normative discussion unless one is willing to pay the price for doing so, and thus if one is going to be inconsistent by the lights of others (which may not be inconsistent in a content-neutral way, but only when considered in light of specific content-specific agreements about what “we” will accept) one must be prepared to pay the pragmatic price for doing so.\textsuperscript{180} Of course sometimes the price will be very small indeed. What a tribesman in Botswana thinks about embryonic stem cell research need not enter into normative discussion for most people living in the United States, since his cooperation in this matter is not something any of us need. The normative issues tied to slavery will never have costs that small, for, of course, at least the slaves themselves will resent their treatment and will be disposed to loaf, rebel, give hateful looks, and such—and these will be costly for the slaveholders. (Perhaps the work would get done more quickly in a more cooperative agreement; perhaps good numbers of slaveholders are killed in a rebellion and then enslaved themselves.) In excluding the slaves from normative discussion as much as they can, the slaveholders pay some real prices—prices with teeth. Moreover, the plight of the slaves may give rise to sympathy on the part of some of the more influential and powerful members of a group whose cooperation in some of the slaveholders’ projects is lost when it could otherwise have been expected (think of wealthy and influential Northern abolitionists and the refusal of Britain to join the Civil War on the side of the South).

In support of the objectivity of the judgment that “slavery is wrong,” Gibbard can appeal to higher-order norms to some degree. As seen above, Gibbard recognizes that there are pressures to be consistent in normative discussion, pressures that—if they were to

\textsuperscript{180} See pp. 160-61, 169-72, and 181-82 above.
include the people enslaved—would push people strongly away from accepting slavery in normative discussion. In the same vein, Gibbard can appeal to evolved tendencies to be impressed with the pragmatic success of science and a drive to be consistent in one’s method of inquiry and with well-established conclusions in science. This will rule out things like supporting slavery by an appeal to phrenology or arguments for Aryan biological superiority. Insofar as these are the grounds of a slaveholder’s judgment that slavery is permissible, there are evolutionarily grounded higher-order norms that militate against endorsing slavery. On the basis of his evolutionary psychology, Gibbard can argue that we face pragmatic pressures toward consistency and toward dropping factual claims that don’t square with a scientific program of inquiry widely acknowledged as wildly pragmatically successful.

Moreover, Gibbard can point out that the real, pragmatic costs of holding a view that slavery is right cannot be avoided altogether in the same way one can easily avoid paying a price for not considering what a Botswanan thinks about the practice of stem cell research in the United States. These costs really do pressure one away from a normative judgment that “slavery is right” and toward accepting a system of norms that renders the judgment that “slavery is wrong.” There is thus an unavoidable pragmatic “objectivity” here.

In this context, Gibbard believes he can “save what is clear about ordinary thought about rationality” and the objectivity people understand as essential to their normative judgments (so long as this is emptied of any lingering hankering for “Platonism”).

Gibbard glosses the pretense of interpersonal authority in terms of conversational demands. The authority behind the normative judgment that “slavery is wrong” resides in the

conversational demands of others in normative discussion, and the fact that the very structure of slavery ensures that some others—namely, the slaves—will bring conversational demands to normative discussion that insistently resist normative judgments that “slavery is right” and resent the “inconsistency” of those who accept such a system of norms (where that inconsistency is not content-neutral, but in light of content-specific affirmations that the slaves themselves accept). This authority will always be in place, since it is likely that (many of) the slaves will not accept the system of norms that prescribes their own enslavement and will demand that others not accept it either. Moreover, a certain sympathy that might be thought to be a part of human “nature” due to our evolutionary history, may lead increasing numbers of others to add their conversational demands to the demands the slaves make. The “authority” residing in these demands is “objective” in the sense that it is present whether the slaveholder likes it or not. The “authority” here is inescapable, being part of the pragmatic realities of the situation.

Much the same may be said for the *pretense of independence*. First, the authority is there independently of whether one accepts the system of norms prescribing that “slavery is wrong” or not. Second, one cannot doubt that the slaves and the abolitionists are sincere in their acceptance of a system of norms that prescribes that “slavery is wrong.” From the point of view of their sincere acceptance of the system of norms that they accept, they themselves cannot help but view the system of norms they accept as holding independently of themselves. That is part and parcel of what it *is* to accept a system of norms. It is a performative necessity that the person who sincerely accepts a particular system of norms think that that is the thing to do, and that others should do so as well. Again, the slaves themselves certainly view the system of norms they accept as giving the thing to do and that
this holds independently of their own act of acceptance. The normative judgment that “slavery is wrong” thus has a certain kind of “independence.”

Finally, Gibbard can also claim that when pressed with regard to the pretense of non-arbitrary reasons, norm-expressivism underwrites a suitable “objectivity” for the normative judgment that “slavery is wrong.” Recall that in the pretense of non-arbitrary reasons, “we seem to distinguish between accepting something as a demand of rationality and making an idiosyncratic existential commitment to it.”

182 Gibbard can appeal to matters of consistency and of the pragmatic difficulty of simply dismissing someone or a group of people as conversation partners in normative discussion to vouchsafe a meaningful objectivity for the normative judgment that “slavery is wrong.” One cannot make an idiosyncratic commitment to just any system of norms because other people will not acquiesce and one will “risk ostracism” and “pay reproductively” for shirking the cooperative demands of social life.

183 One will not be able to pretend that just any “reason” is a reason that others recognize, and there are constraints on what will be able to pass as “rational.” Although content-neutral “higher order norms” do not themselves have the ability to underwrite an objectivity of reasons, the higher order norms enter back into the picture through our evolutionary psychology and the practical demands of life together. “In most cases, then, dogmatic conversational demands are self-defeating.”

184 Not just any act of accepting a system of norms will be accepted by other people, and so not just any “reason” one expresses will count as such in normative discussion with others. There is good reason to


believe that in normative discussion about slavery, the “reasons” offered for “slavery is right” will be pragmatically squeezed out and will be viewed by others as arbitrary and not worth holding. Stripped of conversation partners, the “reasons” one gives oneself in soliloquy will start to seem arbitrary and flimsy even to oneself—not to mention increasingly costly. The “reasons” holding sway around oneself will start to seem increasingly “rational” and one will feel the objective, pragmatic weight of the normative judgment that “slavery is wrong.” As Gibbard puts it, “always having to think secretly is felt as a strain; the person who must do so will soon feel that he is losing his normative moorings.” 185 Thus, because of our evolutionary history, there is “some tendency to gravitate toward the norms of those around one, together with some firmness in sticking to the norms one has hitherto accepted.” 186 In the case of slavery, these tendencies will lend a pragmatic objectivity to the “reasons” for a system of norms that renders the normative judgment that “slavery is wrong.” This is how the psychology will work in regards to us as we consider the morality of slavery.

This gives us all the “objectivity” we could need for the normative judgment that “slavery is wrong.” Admittedly, it is not a “grandiose objectivity” that encompasses all conceivable rational beings, nor even all of humanity. “Grandiose objectivity, though, may not be vastly important,” for it is orthogonal to our everyday lives. 187 “What matters chiefly is not what we can say to strange beings who are merely conceivable, but what we can say to

185 Gibbard, Wise Choices, Apt Feelings, 74.
186 Gibbard, Wise Choices, Apt Feelings, 78.
each other.” In this “modest objectivity,” we recognize that “parochial” judgments are all we need: “Let me call such a judgment—a judgment demanded of a group smaller than all conceivable rational beings—parochial to that group.” We can try to broaden the boundaries of the parish continually, but we need not be concerned if some strange beings could be conceived who meet all the standards of ideal rationality and yet fail to accept a system of norms prescribing that “slavery is wrong.” It is enough that “we” accept such a system of norms. This is all the objectivity we could need, and all the objectivity we should want, for the normative judgment that “slavery is wrong.”

This is, I think, a vigorous statement of a case Gibbard could make for the objectivity of the normative judgment that “slavery is wrong.” What are we to make of Gibbard’s arguments that norm-expressivism satisfactorily secures enough objectivity for these normative judgments for our purposes as human beings who must live together? I think we should not be moved very much by Gibbard’s efforts.

The first thing to note is that pragmatic pressures exerted by one group or another are just that, pragmatic pressures; and these are quite different than normative reasons. Such pressures are not responsive to normative reasons or to rational properties that may be treated descriptively. Indeed, according to Gibbard’s “Galilean” materialism, such things do not exist. Instead, the “reasons” and the “rationality” arise from out of the pragmatic pressures. And so we find Gibbard talking about “costs,” “prices,” and “paying reproductively.” The key thing to note, however, is that these pragmatic costs need not square with any particular normative judgments. They will attend any position that is

---


accepted widely enough, whatever that may be. Looking at the world, both through history and today, one might not be particularly sanguine that these pressures will line up with what should be the case. Slavery has often been practiced through history and continues to be practiced today, and there have undoubtedly been people where the predominance of the pragmatic pressures they faced was toward accepting a system of norms that prescribed slavery for others. And some who resisted such pressures paid a steep price to do so. As it happens, for “us” here in the United States in the 21st century, we face pragmatic pressures to accept a system of norms that prescribes that “slavery is wrong” (though perhaps the majority do not accept norms that prescribe that “extremely low wages for others is wrong”). For many people that has not been the case. Is there a sense in which they should buck the pressures and pay the costs to resist slavery? A full-information/ideal rationality descriptivism or an account of the actual existence of authoritatively prescriptive moral reasons might provide an answer. But, as seen above, Gibbard rejects these paths. True to his “Galilean” commitments, Gibbard holds that the costs with teeth are the pragmatic ones that exercise sway over our non-cognitive acts of acceptance or endorsement.

But what about the use that Gibbard can make of higher order norms, such as demands for consistency and having one’s facts right? It must be noted here again that Gibbard himself does not give higher-order norms a fundamental place in norm-expressivism. To do so would make his theory a version of the descriptivism he roundly rejects. Just as there can be an ideally coherent anorexic or Caligula that fall afoul of no content-neutral higher order norms, there can be ideally coherent slaveholders. Yet, as seen above, Gibbard does acknowledge some place for higher order norms. Admitting that this doesn’t get expressivists all the way to an objective moral judgment that “slavery is wrong,”
how far down the road does it get them? Perhaps this is far enough to gain some significant objectivity.

The first thing to recall in this connection is that these pressures to consistency and getting facts straight owe to evolutionary contingencies: “If content-neutral qualifications rule out wild judgments, they do so in virtue of contingent limits on the ways people are constructed.” Fair enough. It may be that many or most human beings happen to have evolved so as respond (somewhat unevenly) to pressures to form beliefs consistent with other things they have accepted as true and with tendencies to be impressed with the pragmatic successes of methodical inquiry and to subject many beliefs to whatever can be known in this way. But such contingencies will vary somewhat from person to person, and they may be counterbalanced by other contingencies of our nature. And without some more robust account of their grounds, it is difficult to see what a human being who has been shaped somewhat differently by these forces has wrong. Additionally, the contingent limits, seen to be such by the materialist, lose some of their power to bind a person as a person. They are not independent, authoritatively prescriptive moral reasons, but ways that some humans happen to have had evolutionary forces play out.

Let’s look at these issues more closely and apply them to the case of the slaveholder. First, Gibbard has already been seen to allow that at least some (how many?) slaveholders will not be open to criticism on content-neutral grounds. They have taken account of the facts and have a consistent system of beliefs. Insofar as anyone approaches “ideal consistency,” the slaveholder may approximate this ideal. For such people, attempts to correct their views by appeals to such matters will gain no traction. Moreover, given

190 See the full passage at note 109 above.
Gibbard’s views, it is likely that many attempts to appeal to content-neutral standards will amount to little more than browbeating. The slaveholder is consistent by the standards of the people whose cooperation he happens to need for the success of his projects. The contingencies of the culturally formed intellectual starting points will shape raw evolutionary contingencies as they happen to have arrayed themselves in this particular human being. What the abolitionist takes to be an impeccable application of content-neutral higher order norms, the slaveholder takes to be obviously shaped by biases. A more robust full rationality/ideal coherence descriptivist might have a reply that could be appealed to above the contingencies. But Gibbard will only allow that the higher order norms are contingent limits, and he cannot avail himself of the same resources without compromising his norm-expressivism. Gibbard recognizes this and does not shy away from the implication of his expressivism: “We can imagine someone, though, who admits that response R was evoked in conditions C, but who disagrees with judgment J and disagrees coherently. He makes no mistake of language or logic. Rather, he has a different account of what conditions for normative judgment are ideally favorable.”  

Setting higher-order norms is itself a question of what one will endorse, and one cannot go on forever postulating yet another higher-order norm. At some point there must be a substantive stopping point. For the materialist norm-expressivist, this stopping point is endorsement as shaped by evolution and culture. In short, then, given Gibbard’s norm-expressivism, the content neutral pressures are limited by the ways in which evolution has worked itself out in a particular person and are only as good as the group one is a part of. Given the way these evolutionary and cultural contingencies

---

play out, many slaveholders may faces little “content-neutral” pressures to judge that slavery is wrong. Indeed, in at least some cases, it will be the abolitionist who faces such pressures.

Things get worse. Gibbard is clear that higher order norms are norms that govern the acceptance of norms. We have just been reminded that for the materialist norm-expressivist, these norms gain sway by acts of endorsement beholden to evolutionary and cultural forces. The norm-expressivist cannot accept that there are really norms of rationality that can be described properly and adhered to correctly. It’s not as if these norms exist to be related to in this fashion. Insofar as Gibbard really moves in this direction, he gives up his non-cognitivism, his expressivism, and becomes a descriptivist. Rather, it’s a matter of how things fall out. Now, if perhaps these contingent forces will array themselves to spur acceptance of higher order norms of consistency and the like, by the same token they may well take other forms. It could well be that alternative higher order norms for the acceptance of other norms might have more force in light of evolutionary contingencies. Imagine the following higher order norm for acceptance of other norms: accept whatever helps one and one’s progeny be well-placed in society to have one’s projects succeed. Indeed, do we have to imagine that? Since the force of higher order norms arises “in virtue of contingent limits on the ways people are constructed,” it’s at least possible that such a norm for the acceptance of norms might override other possible higher order norms. If the contingencies do line up in this way in a particular person or group of people, what can a norm-expressivist (who is not a full-information/ideal rationality descriptivist) say they get wrong? Only what “we” don’t endorse.

Now clearly some higher order norms shaped by evolutionary and cultural contingencies could favor slavery. Still more, one can know all the facts and simply not be
moved by them. Gibbard, consistently with his expressivism, insists on this as a possibility. I suggest it is an everyday reality. I may be aware that my coworkers are my equals and so on, but none of that may move me in the slightest not to backstab, undermine, etc. insofar as it helps me clamber up the ladder. Gibbard insists that what is missing is the element of endorsement (as seen above). It may be harder for the people reading this dissertation to imagine the same is true with slavery because it is farther removed from our daily experience and is not a live option. But the cases are not really all that different. Awareness of the facts a materialist will countenance as such may remove some ex post facto justifications (like phrenology), but not the endorsement that drives a person to look for such justifications, accept them readily, and devise new ones so far as their psyche needs them (as some may not). Full and vivid awareness of facts may make things a bit more difficult for some views, but the difficulties should not be overstated. As pointed out above, the materialist norm expressivist cannot limit the play of the contingencies by appeal to really existent norms of rationality in a descriptivist manner. And what will be recognized as a fact will vary somewhat depending on how evolutionary and cultural contingencies play out. A quick glance at the human track record on slavery suffices to raise serious questions about whether the “crooked timber of humanity” will underwrite the moral judgment it should. Slavery is morally wrong, but the materialist norm-expressivist is still struggling to show us how his theory can conclude this with any meaningful objectivity.

I conclude that the materialist norm expressivist gets little, if any, mileage from higher order norms. But what about the more directly pragmatic case Gibbard might make? Perhaps that will work where Gibbard’s tepid appeal to higher-order norms does not. We have seen that Gibbard has a more specific case that he can make for the objectivity of the
normative judgment that “slavery is wrong.” The price to be paid by slaveholders will inevitably be steep, including the costs paid in continually exerting energy to oppress those enslaved, the fear of revolt, and the conversational pressures exerted both by (some of) the people enslaved and those (perhaps well-placed and powerful) people who come to sympathize with their plight. Here we have costs with teeth, to be sure; but what if the slaveholders believe that the prices can be managed? It’s not clear that we could know that they must necessarily be incorrect in their calculations of the pragmatic considerations. (To take a contemporary case, maybe the authorities can be counted on to overlook the young girls at the brothel so long as they have complimentary access.) In any case, numbers of people have apparently thought that it was a pragmatically workable situation and that others could viably be enslaved. And at least some of those societies (Greece, Egypt, England, to name a few) enjoyed great ascendancy during the time they practiced slavery. “We” can still say that it is wrong; but it seems that “they” certainly did not need to do so—at least not if pragmatic costs are the issue. Gibbard would certainly be right to say that there are unavoidable pragmatics costs to slavery; what he needs, though, is an argument that such costs always predominate and that people who calculate otherwise inevitably make mistakes in their calculations of the costs. This would not be an easy argument either to make or for most people to follow. At any rate, many societies (or segments of societies), past and present, have tallied up the pragmatic costs differently than “we” do.

Independently of such questions about the how a society as a whole ought to tally up the pragmatic costs, it is clear that the individuals within many societies have faced “conversational pressures” to accept a system of norms prescribing slavery. For these individuals within such societies, the “conversational pressures” at the heart of Gibbard’s
analysis of the *pretense of authority* actually favor slavery. In some of these societies it could be the case that one could buck those pressures only at great cost to oneself, including the failure of projects that would otherwise have succeeded and in terms of less reproductive success. It’s difficult to see how Gibbard could avoid the conclusion that the normative judgment that “slavery is right” would have “authority” for that individual within that society, given his analysis of authority in terms of conversational pressure. Indeed, it’s likely that many people in history and still today have been in this situation. Gibbard could, of course, avoid this difficulty by abandoning his materialist non-cognitivism in favor of a full-information descriptivism or a theistic objectivism or perhaps by pursuing some other tack. But, to date, he has steadfastly refused to do this. Thus he faces the basic problem that the pragmatic pressures, including the conversational demands made by various parties in normative discussion, can be exerted on behalf of nearly any stance. Thus, which normative judgments are “authoritative” appears to hinge on a question of which group’s cooperation you need for your projects to succeed.

Perhaps the most basic problem, however, is that it seems that where what was needed was an authority residing in or somehow uniquely attached to the normative judgment that “slavery is wrong,” what Gibbard has given us is a welter of competing authorities—with it being a contingent matter of the accidents of birth and upbringing which normative judgments will happen to be authoritative. Where we need an authority, Gibbard gives us authorities; where we need objectivity for normative judgments, Gibbard gives us objectivities. This clearly creates problems for Gibbard’s account of the independence of normative judgments as well. Undoubtedly many people have sincerely accepted all sorts of systems of norms, including that slavery is right. On Gibbard’s account of independence,
each person who truly accepts any such system of norms thereby believes that it is valid independently of that act of acceptance. Returning to the case of the “ideally coherent anorexic,” a person’s sincere acceptance of a system of norms that prescribes that one not starve oneself for the sake of a trim figure means that in “a hypothetical case in which we ourselves are ideally coherent anorexics, these norms prohibit us too from starving. It is in this sense that we think that the prohibition against starving for a trim figure is valid independently of our own acceptance of it.”192 The problem is that any number of people can claim the same “independence” for any system of norms they sincerely accept, including the “ideally coherent anorexic” and the person who sincerely accepts a system of norms prescribing slavery. My sincere normative judgments get to claim “independence,” but so do the incompatible systems of norms sincerely accepted by others.

Much the same sort of problems beset Gibbard’s attempt to do justice to the pretense of non-arbitrary reasons. First, the loss of conversation partners and the subsequent loss of orientation that anchors one’s reason can cut in many different directions. It could be the one who accepts a system of norms prescribing slavery who finds himself cut off from normative conversation and the benefits of social cooperation; but it could also be the one who rejects such a system who finds the norms she accepts regarded as arbitrary and who has no conversation partners to normalize her “reasons.” Second, whatever norms a self-aware norm-expressivist accepts, she will realize that she is responding to pragmatic pressures, sincere though she may be, and that others are accepting different systems of norms on the same basis. Competing systems of norms share the same pragmatic ground, modified only by the contingencies of one’s situation. Moreover, she will realize that she too would likely

192 Gibbard, Wise Choices, Apt Feelings, 166.
have had the same system of norms as those others had she been in their situation, for in
that case she would have been responding to the pragmatic costs that would have had teeth
for her had she been in their place. As a slaveholder, she might see the arbitrariness of her
reasons. She might also see the same thing as an abolitionist. In each case, she might well
continue to believe she is right to accept the system of norms she accepts given who she is,
but she might also see the arbitrariness of her reasons—especially if, as a norm-expressivist,
she believes that there are no normative or rational facts of the matter.

Perhaps none of this should concern us very much, however. Isn’t all this a chasing
after a “grandiose objectivity,” when a “modest objectivity” will do? Gibbard admits that a
kind of “parochialism” will attend his norm-expressivism, and so perhaps the norm-
expressivist can simply dismiss all these worries with a wave of the hand or a shrug. Why
worry about a few oddballs who have an unusual psychological makeup or are set in bizarre
cultural situations so that they endorse some outré views? Surely this doesn’t affect the
morality of the vast majority of people. But in what sense are we talking about “unusual
psychological makeups”? Statistically? Across all cultures? Across all times? Is the
psychological disposition to slavery odd, in those senses? Perhaps it is the one who takes a
costly, principled stand against slavery who has an “unusual psychological make-up.” Again,
a look at the “track record” for human beings on the issue of slavery would indicate that the
“normal psychological makeup” isn’t much to admire. Or, perhaps, it’s odd in relation to
what “we” think? But who is this “we”? Whence its normative privilege, given materialist
norm-expressivism? An “objectivity” limited to “us” is very modest indeed. Gilbert
Harman could give us that.
Moreover, it is simply not true that the worries about objectivity are far removed from the daily moral decisions we must make. Far from it. From abortion to sweatshops to sexual slavery to nuclear arms to domestic violence (to name a few), the objectivity in view here is very much the stuff of lived life. In all these situations, the objectivity of interpersonal authority, independence, and non-arbitrary reasons is the stuff of regular moral thought and discussion. And people regularly claim objective backing or support (authoritatively prescriptive moral reasons) for norms they recognize. As Gibbard recognizes, “Any account of [our] language that ignores this claim must be defective. It may capture all that the speaker could claim without illusion, but it will not capture all that he in fact is claiming. It will not be a genuine analysis of the concepts involved, but a way of defanging a claim that originally had bite.”\(^{193}\) One need not be engaged in a project of “grandiose objectivity” disconnected from real life (if such a project would be) in order to argue that the normative “objectivity” of Gibbard’s analysis goes beyond modest to meager.

In the end, Gibbard’s “Galilean” norm-expressivism does not sustain a recognizable objectivity for normative judgments, including clear moral cases where it should: the severe difficulty he has in underwriting the objectivity of the normative judgment that “slavery is wrong” is a case in point. The central problem is that where we needed normativity we were given pragmatic pressures, and that Gibbard’s analysis of the key notions of objectivity in these pragmatic terms that fit his “Galilean” materialism gives us the “botched likeness” he was worried about.\(^{194}\) Where we needed an authority uniquely tied to a normative judgment, we received many authorities pointing in different directions, each tied to the


\(^{194}\) See the epigraph to this dissertation.
“conversational demands” of different groups, none of which are responding to normative facts or rational-making properties of the judgments themselves (these not existing in the “Galilean” world). Where we needed normative judgments that could be seen to hold independently of our acceptance of them, we find only a highly attenuated sense of independence tied to the very fact of that acceptance itself—a sense of independence we realize that those who sincerely accept divergent systems of norms can also help themselves to. And where we needed non-arbitrary moral reasons, not tied to the peculiarities or contingencies of one’s situation, we are left with moral “reasons” that express sincere acceptance of a system of norms—an acceptance tied to the group whose cooperation one happens to need if one’s projects are to succeed and if one is to avoid having to “pay reproductively.” In short, Gibbard has found his way into the “coal pit” of moral reasons arbitrariness.

At the end of reading Gibbard’s account of human normativity in Wise Choices, Apt Feelings, one gets the feeling that the subject has been changed and that some critically important matters have dropped out of the new subject. Gibbard’s new subject fails to capture the essential matter at stake in each of the three “objective pretensions” of our normative life, doing justice neither to the independence of normative judgments, nor to their authority, nor to their being a demand of rationality instead of accidents arising from the way contingent matters happened to fall out. The central problem throughout has been Gibbard’s failure to give any robust notion of moral reasons and their role in moral discussion. At the heart of Gibbard’s understanding of normative reasons is a reduction of these reasons to a matter of acceptance that opens his views up to a host of very troubling difficulties. Normative reasons, being a matter of expressed endorsement of a system
prescribing particular behaviors, end up hostage to the forces acting on the particular person
who does the accepting. The various particular endorsements of this or that system of
norms are pitted against each other, and normative reasons neither limit what is accepted
(reasons themselves merely expressing precisely that acceptance) nor adjudicate the conflict
of endorsements (since they merely restate the conflict). Normative reasons as expressions
of endorsement turn out to be in hock to a host of adventitious forces, particularly of an
evolutionary and cultural sort. The upshot is a picture of the normative life where the real
action has to do not with entering into earnest discussion where we are responding
normative reasons as such, but with conversational demands and other pragmatic concerns.
Independent, authoritatively prescriptive reasons have dropped from view. While there is a
kind of prescriptivity addressed to people in the “conversational demands” of others, it is
neither independent nor authoritative. Gibbard’s norm-expressivism founders on a moral
reasons arbitrariness for which he has no apparent solution—an arbitrariness that is only
exacerbated by Gibbard’s appeal to pragmatic concerns alien to normativity as such.

Gibbard’s more recent work does not correct these fundamental defects in norm
expressivism. As we turn to Gibbard’s subtle and difficult 2003 book, Thinking How to Live,
and then compare Gibbard’s work to that of Stephen Darwall, it will become clear that
Gibbard continues to lack an account of authoritatively prescriptive moral reasons that
would allow him to avoid the problem of moral reasons arbitrariness.
II. Thinking How to Live

A. Plans, Properties, and Concepts

Toward the beginning of his book, *Thinking How to Live*, Gibbard states his programmatic slogan: “Thinking what I ought to do amounts to deciding what to do.”\(^{195}\) He immediately flags an important objection, noting that if the views expressed in his slogan are right one might wonder if people can “ever be mistaken in an ought judgment.”\(^{196}\) It is not without reason that many have pressed this objection against Gibbard’s norm expressivism, and in *Thinking How to Live*, Gibbard sets his sights on a convincing reply: “This kind of objection spurs much of what I do in this book.”\(^{197}\) He hopes to show that “even when we make no mistakes about those natural facts [with which our normative expressions have to do], we can make wrong decisions.”\(^{198}\) In this sense, *Thinking How to Live* is an extended defense of the ability of norm expressivism to give a convincing account of a central “objective pretension” of our human normative life. This section of the dissertation will examine these claims carefully and argue that Gibbard still has not captured the objectivity of normative judgments.

Much of *Thinking How to Live* deals with an issue that is not the focus of this dissertation, namely, working out a convincing reply to the embedding problem first raised by Frege and made famous by Geach. Gibbard seems to fall well short of a satisfactory

---

\(^{195}\) Gibbard, *Thinking How to Live*, 17. Gibbard first announces this slogan on pages ix-x. This slogan also appears in Gibbard, *Reconciling Our Aims*, 19.

\(^{196}\) Gibbard, *Thinking How to Live*, 17

\(^{197}\) Gibbard, *Thinking How to Live*, 17

\(^{198}\) Gibbard, *Thinking How to Live*, 17
reply to this problem, but nothing more will be said of it here.\textsuperscript{199} Instead, for the sake of argument Gibbard will be granted his reply to the thorny logical and semantic issues raised by the embedding problem in order to focus attention on his continuing inability to give a convincing vindication of the “objective pretensions” of our human normative life. In short, Gibbard remains mired in moral reasons arbitrariness and the problems attending it.

The core notions of Gibbard’s norm expressivism remain largely unchanged over the thirteen years between \textit{Wise Choices, Apt Feelings} and \textit{Thinking How to Live}. There are, however, some modifications. An obvious change in Gibbard’s later statement of norm expressivism is that the notion of a \textit{plan} has taken center stage. Gibbard states that he thinks of his “shift from norm-acceptance to planning not as a change of position but as a shift of expository purposes,” and this seems to be an accurate assessment.\textsuperscript{200} The change of expression allows Gibbard to distance himself from the idea of accepting a \textit{system} of norms, which cannot be readily assimilated by norm expressivism because of the obvious reliance on notions of consistency and inconsistency. Additionally, it serves to emphasize that normativity arises from human activity, from a \textit{state of mind} rather than “seeming beliefs,” thereby removing “the need for non-naturalistic mumbo-jumbo.”\textsuperscript{201} There is nothing to our seeming normative beliefs other than the conceptual commitments incurred in our activity of planning. Finally, Gibbard employs the idea of planning in developing an apparatus for capturing the quasi-content of expressed commitments that mimics a possible


\textsuperscript{200} Gibbard, \textit{Thinking How to Live}, 181.

\textsuperscript{201} Gibbard, \textit{Thinking How to Live}, 192. “The expressivist starts with states of mind, and uses these to elucidate normative beliefs or seeming beliefs” (180).
This is part and parcel of his attempt to defuse the embedding problem. “Plans,” Gibbard argues, “are judgments, in that they can act in many ways like beliefs in plain fact. They combine with beliefs in all the ways that beliefs can combine with each other. Standard logic applies, explained in a way that also explains the logic of belief.” In *Thinking How to Live* and in his recent Tanner Lectures, *Reconciling Our Aims*, planning replaces acceptance as the central human activity giving rise to our normative life, and it does so in ways that capture all the familiar logic of our normative reasoning—or so Gibbard argues.

As his “canonical” form in which our planning activity comes to expression, Gibbard gives us the following: “φ-ing now is the thing to do.” It is of critical importance that this is understood not as stating a belief but as bringing the activity itself to expression. Gibbard places particular emphasis on this point: “what’s the psychological difference between planning to pack and believing that one so plans? The latter is a psychological belief, whereas the former is no prosaic belief at all, but a state of planning—

---

202 Gibbard speaks of “hyperplans” that are “much like the possible worlds of the possible worlds semanticist” (Gibbard, *Thinking How to Live*, 57).

203 Gibbard, *Thinking How to Live*, 137. As the passage states, Gibbard reduces logic generally to this expressivist model.

204 Gibbard, *Thinking How to Live*, 41. As with his confession in *Wise Choices, Apt Feelings* that he was attempting to “exploit” a meaning, rather than capture ordinary usage, Gibbard here also works by fiat: “By sheer stipulation, then, the meaning of this phrase ‘the thing to do’ is explained expressively” (8).

205 Philip Pettit summarizes the idea nicely: “the idea is that the thinking and deciding are not ‘separate activities’. In particular, the thinking does not count as an instance of representing, only as a case of deciding or planning.” (Philip Pettit, “On Thinking How to Live: A Cognitivist View,” *Mind* 115 (October 2006): 1091). Pettit notes that Gibbard insists that the phrase “thinking what to do” has it expressivist meaning only as a matter of stipulation. Pettit levels strong criticism at Gibbard on this point, arguing that Gibbard fails to avoid falling back on a “presumptive rationality”: “It may seem, of course, that in order for me to pick out an alternative as ‘the thing to do’, holding or making a plan to do it, I must form a representation of it as the thing to do—a representation expressed by saying it is the thing to do” (1092). Although Pettit’s criticism is not the criticism being pursued here (though not wholly unrelated), his criticism is well-taken. However, Pettit doesn’t seem to see as clearly as Gibbard does the difficulties that that Pettit’s strong cognitivist understanding of belief raises for naturalism.
that’s the contrast that lies at the heart of expressivism.”\textsuperscript{206} When I say, “Writing now is the thing to do”, it is \textit{my activity} of planning itself that is being \textit{expressed}; there is no belief being stated such that I am \textit{saying that} some \textit{state of affairs} is the case. All such non-materialist mumbo-jumbo is kept safely at bay in favor of psychological states readily amenable to materialist explanation.

With this critical distinction in place, Gibbard asks us to perform a thought experiment in which we imagine a supercompetent planner whom he dubs “hyperdecided Hera.”\textsuperscript{207} Hera forms a “hyperplan” that not only takes all facts into account, but decides in advance what the thing to do is in any possible situation in which she might conceivably find herself. “Every act open on every possible occasion for choice she either permits or rules out.”\textsuperscript{208} And it is important to note that there can be \textit{multiple, equally competent hyperdecided planners}. “If hyperdecided Zeus and Hera agree on all matters of prosaic fact but not on how to live, they will agree whether act \textit{a} is \textit{p}-okay [okay with respect to one plan or another], but may disagree whether it is okay.”\textsuperscript{209} In other words, Hera and Zeus agree on the facts, but disagree about “the thing to do” given a particular situation with respect those facts. Nevertheless, Zeus could see that within Hera’s hyperplan \textit{p}-ing would be the thing to do, even though he believes she is mistaken. The “disagreement” here does not have to do with the facts—by definition, Hera and Zeus agree about the facts—but with how one

\textsuperscript{206} Gibbard, \textit{Thinking How to Live}, 76.

\textsuperscript{207} Gibbard’s thoughts here seem to be a development of his thought extending back at least as far as, \textit{Wise Choices, Apt Feelings}. See pp. 94-97 of that earlier work.

\textsuperscript{208} Gibbard, \textit{Thinking How to Live}, 90.

\textsuperscript{209} Gibbard, \textit{Thinking How to Live}, 93. “Given any hyperplan \textit{p}, an act is \textit{p}-okay in a situation or not—and whether it is is a matter of prosaic fact. An act is \textit{p}-okay in a situation just in case plan \textit{p} permits that act in that situation” (89).
should engage those facts. The difficulty is that “their hyperplans are at odds. Hence the explanations that the two think best, explanations of how to live and why, are likewise at odds.”

Or, as Gibbard states the matter a little later: “their disagreement is purely practical; it is a difference over what ultimately to aim for in life.” This disagreement preoccupies Gibbard throughout *Thinking How to Live*. His central concerns have to do with the logical status of this disagreement when it is conceived in expressivist terms (that is, the embedding problem), and with our ability to carry on such disagreements in a quasi-objective manner.

Before critically examining Gibbard’s best efforts to underwrite some sort of objectivity for disagreements in plan, we need to elucidate one more distinction that Gibbard leans on heavily throughout *Thinking How to Live*. Gibbard, consistent with his aim of cleansing philosophy of any non-materialist mumbo-jumbo, insists that there are no normative properties. No “queer” properties are lurking in corners of the universe; “there are just plain old properties and relations.” These properties have to do with the facts of which Hera and Zeus are aware, and they are purely descriptive (that is, materialist). One must distinguish, however, between properties and concepts, and some concepts can be what Gibbard calls “plan-laden.” While all properties are materialist, some concepts can be “non-naturalistic.” Why? Because they result from our planning activity. In “plan-laden”

---


211 Gibbard, *Thinking How to Live*, 223; emphasis in original. Recall here the “ideally coherent anorexic” from *Wise Choices, Apt Feelings*.


213 As Gibbard puts it, “there’s no such thing as a non-descriptive property” (*Thinking How to Live*, 115).

214 Gibbard, *Thinking How to Live*, 7, 32, 98, 114. Gibbard sees his own view as having affinities with that of G.E. Moore at this point, although he takes Moore’s open question argument as showing a “distinctness of concepts, not of properties” (32).
concepts we brush “to be doneness” onto a materialist world where normativity never inheres in properties. “Some concepts are specially plan-laden; the concept of being okay to do is a prime example. But no property is plan-laden.” My plan-laden concepts arise from planning to live in a particular way, and in this way a non-naturalistic “to be doneness” shapes my engagement with the factual world with which I have to do. Thus, I may have a descriptive concept of tending-toward-pleasure, but there is a separate, plan-laden concept of being the thing to do. Applying these ideas to Hera, we see that Hera may have a descriptive concept of tending-toward-pleasure, and—because of her planning to live in a certain way—may apply the separate, plan-laden concept of “to be doneness” to that descriptive concept, so that Hera thinks that the thing that tends toward pleasure is always the thing to be done. In Zeus, however, these come apart, so that Zeus joins “to be doneness” to the concept of tending-toward-the-perfection-of-one’s-abilities in ways that sometimes do not tend toward pleasure.

Gibbard understands plan-laden concepts (not properties) to supervene on materialist properties, and a major question his views give rise to is how to understand the relationship between the subvenient properties and the supervenient concepts—especially given their disparate provenance. Typically one thinks of supervenience as connecting properties, as between something’s having the property of shattering when a specific amount of pressure is applied and its having such and such molecular properties, and for this reason Gibbard is conceded that his view here might best be called “quasi-supervenience.” The difficulty
here is that there is an intervening layer—namely the planning activity itself—that threatens to pull the (quasi-)subvenient properties apart from the (quasi-)supervenient plan-laden concepts. What would appear to be needed in order to glue the concepts to the properties in a quasi-supervenient manner is a psychological story that itself supervenes in the strict sense on materialist properties, where this psychological story accounts for our planning activity. Gibbard’s speculative evolutionary story in *Wise Choices, Apt Feelings* might fill this need, but some of the difficulties with that attempt have already been seen in the previous section. All this raises questions of how the plan-laden concepts are related to the materialist properties. Nevertheless, Gibbard declares that someone who makes plans, “as a planner, is committed to thinking that there *is* some property that constitutes being okay to do.” Any story Gibbard could tell here would be an evolutionary story that faces

---

218 In *Thinking How to Live*, Gibbard has a curious chapter on supervenience. In the final analysis he is unwilling to allow that plans/rationality/normativity involve a property that supervenes on the factual world. Rather, as far as I can tell, Gibbard does not even allow “normative properties” that have the dignity of supervening on something that is “really” going on. They are just eliminated. For Gibbard, it’s “concepts” that are doing the heavy lifting: “Properties are just properties, neither descriptive or plan-laden. Some concepts of a property, though, are descriptive, and other concepts of the same property are plan-laden” (98). The psychological states that underlie such conceptions must, however, supervene on the material stuff in some way, though, it would seem. Gibbard does at one point say that “there is a factual property that constitutes being okay to do” (97), characterizing this as the conclusion of a transcendental argument; but he says that “this is no real constraint at all… any way of living satisfies it” (97-8). But Gibbard is being quite slippery here again. He is not saying that there is any normative property that does in fact supervene on some subvenient property, but that any planning agent must act *as if* there is such a property. This is, as he calls it, a “quasi-supervenience” (94). He thus does not commit himself to any actual normative properties, but is able to help himself to there being a performative necessity to think as if they did in fact exist. It is in these performatively necessary conceptions that Gibbard’s quasi-objectivity is supposed to reside, and it is here that he makes his attempts to salvage something of the “objective pretensions” of our normative activity. All objectivity is an “as if” objectivity. Gibbard thus allows no normative *properties* that have any sort of causal efficacy (via supervenience), but insists that no one can avoid taking their own normative *concepts* in such a way that they think of them as having such substance underneath them. Gibbard’s normative expressivism is thus understood by him to be unadulterated by any normative properties whatsoever, even as he lays out his quasi-objectivity. Gibbard here seems to parallel some lines of Kant’s thought in the *Critique of Judgment*, a parallel that was also noted in connection with Gibbard’s earlier work by Simon Blackburn, “Wise Feelings, Apt Reading,” *Ethics* 102 (January 1992): 346. Although exploring these connections would be fascinating, doing so take us too far afield for our purposes here.

219 Gibbard, *Thinking How to Live*, 94-5; emphasis in original.
the difficulties seen above, and any such story will obviously be complex and messy. It is unclear how any such story could solve the kinds of difficulties plaguing Gibbard’s failed attempts to do justice to the “objective pretensions” of morality in *Wise Choices, Apt Feeling*—a failure that mires his project in moral reasons arbitrariness. But I will not repeat the difficulties already canvassed in a new register, for perhaps there are deeper difficulties for Gibbard’s moral philosophy lurking in the area.

Having sketched Gibbard’s complex machinery, let’s recall his main point: there are no normative properties, but only plan-laden concepts resulting from our activity as planners. It is these concepts that constitute our everyday normative experience. Nothing here poses any problems for a materialist explanation, and nothing “spooky” threatens a tidy metaphysical explanation.

**B. Settling the Telos of a Plan within a “Galilean” Metaphysic**

Just beneath the neat surface of Gibbard’s materialist account lurk a number of severe difficulties, and the basic problems are easily grasped. Although there are many such problems, the focus here will be on those that show that Gibbard has not captured any interesting semblance of normative objectivity.

Recall first the expressivist’s burden: “The expressivist starts with states of mind, and uses these to elucidate normative beliefs or seeming beliefs.”\(^\text{220}\) Thus, in Gibbard’s version of expressivism, our seeming normative beliefs are “an intelligible consequence of the nature of planning and deciding.”\(^\text{221}\) This is the human activity at the root of

---

\(^{220}\) Gibbard, *Thinking How to Live*, 180.

\(^{221}\) Gibbard, *Thinking How to Live*, 5.
normativity, and—since all properties are natural properties that are not the right kind of stuff to carry any of the weight of normativity—the weight of all (quasi-)objectivity within this picture must be shouldered by our activity as planners and deciders. In this Gibbard seems right; “materialists” who postulate some sort of *sui generis* non-materialist properties (a la Moore and contemporary metaphysically monist “property dualists”) have a pastiche that illicitly borrows normative capital from another metaphysical outlook. Gibbard chooses the honest toil his metaphysics requires of him. In a world of materialist properties that Gibbard believes (rightly, I think) are the wrong kind of stuff to carry normative implications on their own, normativity originates on the human side. Only here do we find something that might give rise to the non-natural “to be doneness” inherent in normativity, in the form of non-naturalism about some concepts. Gibbard is well aware of the difficulties of biting this bullet, but he sees that the materialist has to bite the bullet nonetheless. Gibbard, as with Jaegwon Kim in the philosophy of mind, would not have his physicalism “on the cheap.”

The problem, however, is that the price is steep. An obvious question is what will constrain our activity as planners, since—as Gibbard rightly sees—the facts of a materialist world impose only the bluntest constraints of what just won’t work. If my plans involve my jumping over a three story building or knocking out the world heavyweight champ with one punch or having everyone in the world pay me 15% of their income, things won’t go well

---

222 Or so I would argue anyway, along with Gibbard at this point. It is not, however, in any way the burden of this dissertation to address ethical non-naturalism or non-reductive materialism.

223 Gibbard, *Thinking How to Live*, 32.
for those plans. But so long as what I propose to do is “viable,” anything goes so far as the materialist properties of the factual world are concerned. If I plan to enslave others or beat my wife, and if doing so is viable, the properties of the physicalist’s natural world will not say otherwise. What, then, will constrain our planning activities (or our endorsement or our “Hooray!” or whatever human activity is taken to give rise to our normative life)? For a materialist, the constraint, Gibbard sees, must be a human constraint—and one that could plausibly have arisen from evolutionary mechanisms. No natural properties will do the trick; so concepts of human provenance must do the trick, somehow. It is here that Gibbard hitches his wagon to plan-laden concepts.

To see the problems in Gibbard’s gambit, it will be helpful to consider the nature of plans. Plans involve at least the following elements: (1) a view of the current factual situation, and thoughts about the various ways in which that factual situation could unfold, (2) the end or goal that would be attained by acting within this unfolding factual situation in a particular way, and (3) thoughts about how to achieve the specified end by engaging whatever particular factual situation actually unfolds in the event. Call these parts, respectively, (1) the factual outlook, (2) the telos of the plan, and (3) the proposed attainment method. This is not how Gibbard lays out the idea of a plan, and in fact he has surprisingly little to say on the matter of what exactly a plan is. In particular, Gibbard doesn’t clearly lay out the second and third elements of a plan, contenting himself with an ambiguous “what to

---

224 There is obviously a distinction in these constraints. With respect to jumping over a building, this is the bluntest or most purely physical constraint. The properties of the natural world simply won’t let me do this. As we move toward the constraints that prevent a situation in which everyone gives me 15% of their income, the constraints are less directly physical. It is these sorts of constraints specifically that Gibbard has in mind when he talks about the pragmatic constraints “with teeth.” On these, see notes 86-88 above.

225 There is no claim that this definition is comprehensive, but plans do involve at least these three elements.
do” or “the thing to do,” which could mean either one’s telos in a plan or the attainment method. It is important, however, to keep these distinct.

Perhaps part of the reason why these distinctions may have been elided in Gibbard’s account is that his theory faces particularly difficult problems with respect to a plan’s telos. Gibbard speaks often of “settling” the thing to do, and it is precisely with respect to settling the telos of a plan that Gibbard faces his most difficult problems—especially with respect to doing some sort of justice to the “objective pretensions” of morality.226 For in a plan’s telos we have a crucial moment of acceptance or endorsement that Gibbard has no obvious way of deciding in a principled manner; the problem is that normative reasons have gone missing from the business of settling the telos. Recall the nub of the conflict between hyperdecided Hera and Zeus: it does not turn on factual matters, but rather “their disagreement is purely practical; it is a difference over what ultimately to aim for in life.”227 At its farthest reaches, the telos of a plan encompasses what to aim for in life—what kind of life to live. But this is not settled for us by the factual world; it is “plan-laden.” Plants and most animals might have the telos of their life settled by factual matters, but we don’t. For Zeus and Hera, then, their differing views about what ultimately to aim for in life fundamentally shape their plans.228 Ex hypothesi, Hera and Zeus agree about the factual outlook and thus agree about the factual situation, S, facing agent, A—and yet they disagree about what is “the thing to do” for A in S. For Hera, pulling the trigger is the thing for A to do in S, while for Zeus not pulling the

226 For the “settling” language, see Gibbard, Thinking How to Live, x, 10, 55, 130, 161, 163 and passim.

227 See note 211 above. I have added the second emphasis.

228 The problem here is what I have elsewhere referred to as the problem of “orientational projectivism.”
trigger is the thing for $A$ to do in $S$. The differences between what Zeus’ hyperplan and Hera’s hyperplan would have $A$ do in $S$ stem from the different ways in which Hera and Zeus have settled the telos of their plans. Furthermore, even when Hera and Zeus agree about the thing for $A$ to do in $S$, their reasons for this are different because the “reasons” for their planning to do one thing rather than another if one is $A$ in $S$ flow out of the different aims enshrined in their plans. Zeus has settled on one telos and Hera on another; accordingly, divergent reasons for action govern life within their plans.

So, how is the telos of a plan settled when “there is such a thing as knowing how things stand in our surroundings, but not such a thing as knowing ultimately what to aim for”? This is a central question for Gibbard’s project, and it is one he does not face as squarely as one would like. This much is clear, however; the factual world cannot directly settle the telos for someone’s plan. “‘Facts’, then, are true thoughts that are not plan-laden. They do include psychological facts of how one does plan, but they don’t include what’s okay to do or not.” Once all the facts are in, there still remains the crucial matter of settling the telos of one’s plan. The facts and the telos must be kept separate. The “full, true story” explaining our normative life “factors into a planning part and a naturalistic part, and the naturalistic part contains the full causal explanation.” At the center of the “planning

---

229 A “hyperplan” is a plan “for every situation one might conceivably be in” (Gibbard, Thinking How to Live, 54). Note also that the “content” of claims about what to do is fleshed out by what “hyperdecided” planner would agree with your plan (205-7).

230 Allan Gibbard, “Précis of Thinking How to Live,” 697.

231 Gibbard, Thinking How to Live, 203.

232 Gibbard, Thinking How to Live, 214.
And here we meet with the central element of Gibbard’s project that makes it an expressivist project. At some point an expressivist must make endorsement central so that a human activity rather than a belief reaching out to a state of affairs shoulders the load in explaining our normative life. In *Thinking How to Live*, endorsement is made central at the point where one plans what to aim for in life. As Gibbard insists, this is the purest of planning activities.

But here the expressivist’s central difficulty rears its head: What reasons could we possibly have for endorsing one thing over another? As in *Wise Choices, Apt Feelings*, Gibbard again rejects full-information/ideal rationality theories. Indeed, in an entire chapter Gibbard says he will “look at ideal response definitions and argue that they fail—at least in their fullest ambitions.” Gibbard’s line of response is familiar: “You and I might, if rendered ideal, make different plans for the same exact plight, the same exact hypothetical contingency.” This, of course, sounds very much like what Gibbard said in *Wise Choices, Apt Feelings*, and his position has largely remained the same. This is seen also in examples like that of Hera and Zeus. Most basically, Gibbard argues that any “ideal response definition” will be “plan-laden,” such that the planning activity must in a fundamental sense precede the identification of the ideal: “Being *ideal* as a responder, I say, is itself a plan-laden

---

233 Gibbard, *Thinking How to Live*, 229; emphasis added. Gibbard speaks of this a “the fundamental question in living” (231). An interesting question is how one might characterize a person’s plan to follow deontological moral principles. Apparently, then, there would be no good that her plan aims at. As a utilitarian, Gibbard might not be overly concerned about this question. Additionally, deontologists like Korsgaard and (perhaps) Kant would see a kind of good operative in deontology after all: perhaps something like the dignity of pure respect for the law itself. In this case, there would be a telos in a deontological plan after all.


concept.” Or again, even more directly, “the concept of an ideal frame of mind for decision seems itself to be normative. Hence we might best treat this concept as plan-laden: an ideal frame of mind, we can say, means one to trust. To regard a frame of mind as ideal for planning is to plan to trust planning done in that frame of mind. Reading the term ‘ideal’ this way eventuates in ideal response definitions that are plan-laden.” The telos of a plan cannot be constrained by some sort of ideal rationality because these plans themselves precede and shape what is recognized as ideally rational.

If Gibbard cannot look to some sort of “ideal rationality” to guide our acceptance of one telos or another, neither can he look for help from the world outside us. Since the materialist finds nothing addressing normative reasons to us from the “Galilean” world of fundamental fact which has undisputed rights to causal explanation, we are cast adrift on the question of what our plans should aim at. “Life faces us with hard questions. I wish that questions of how to live had clear answers (though perhaps they do for the lilies of the field or for giraffes, and I’m not wishing I were one of them).” Gibbard makes it clear in Thinking How to Live (as he had done previously in Wise Choices, Apt Feelings) that the Darwinian mechanisms by which the Galilean world gave rise to us do not speak to us on the matter of how to live. Trees overtop one another and the lead lion kills the cubs of his predecessor upon gaining ascendancy. They are completely submerged in the facticity of their nature; but we, somehow, have gained a measure of distance from the overwhelming

---

236 Gibbard, Thinking How to Live, 237; emphasis in original.
237 Gibbard, “Précis of Thinking How to Live,” 696; emphasis in original.
238 Gibbard, Thinking How to Live, 16-17.
239 Gibbard, Thinking How to Live, 264-5.
crush of facticity. Unlike them, we can choose an end; the difficulty is that in a Galilean world devoid of any fundamental address to us as we go to make that choice it is very difficult to see why choosing to aim at one end should be any better than choosing to aim at another. Reasons for settling the telos of a plan one way rather than another are in very short supply for Gibbard.

At first glance, Gibbard’s dearth of reasons relevant to settling the telos of a plan owes to his expressivism, and this is certainly true. But Gibbard’s norm-expressivism is not the whole problem. In an important passage, Gibbard anticipates the objection that in talking about plans he has simply helped himself to reasons in a way that is illegitimate for the expressivist. When we make a plan and act in accordance with it, after all, we certainly seem to be acting on the basis of reasons—and these seem to be irreducibly normative. It can look very much like the “expressivist helps himself to all that a non-naturalistic realist needs.” In responding to this challenge, Gibbard seeks to defang the notion of reasons, offering the notion of a “plan” as a replacement for reasons. Responding to a challenge from T.M. Scanlon that he had smuggled in the normativity he needed, Gibbard argues for

240 One of the most incisive thinkers on this sort of question is Martin Heidegger. Indeed, there is a real sense in which the question he gave his whole life to (and finally, I believe, failed to answer)—the Seinsfrage—is really the question of how we gained this distance from the crush of facticity. It may be that the reasons behind the lack of an ethics in Heidegger’s thought are similar to the sorts of problems Gibbard encounters. While drawing out this connection would be fascinating and would cast light on just how deep Gibbard’s difficulties run, this dissertation is not the place to pursue them.

241 Note that Gibbard speaks of a “choice of final end” (Gibbard, Thinking How to Live, 230). Perhaps this language of choice overemphasizes the extent to which these things are within our control, but it serves to emphasize one pole that threatens to engulf Gibbard’s project. The problem for Gibbard is that as he pushes off against this pole, the other pole—that of one’s “choice” being determined by one’s evolved nature—threatens from the other side. An important critique of the idea of radical choice in these matters is Philippa Foot, Virtues and Vices (Berkeley, CA: University of California Press, 1978); see also Simon Blackburn, “The Flight to Reality,” in Virtues and Reasons: Philippa Foot and Moral Theory, ed. Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn (Oxford: Clarendon Press, 1995).

242 Gibbard, Thinking How to Live, 189.
“another reading” of reasons: “To ‘treat R as weighing in favor of doing X’ is to plan in a certain way, to figure out what to do in a certain way, a way we can describe without helping ourselves to the notion of being a reason. It is to weigh a consideration R toward doing X. We start, then, with a psychological notion, a person’s weighing a consideration toward taking a course of action.”

Gibbard seeks to replace the normative notion of a reason with a non-normative notion of a plan—the latter being rooted in a strictly psychological notion. If a reason did in fact count in favor of doing something, the materialist would have run into something irreducibly normative and—from the standpoint of her metaphysic—unexpected and queer. And so Gibbard, following the dictates of his metaphysical commitments, eliminates reasons and proposes “plans” as a replacement notion.

There are many important issues here worthy of exploration, but for the purposes of this dissertation, what is crucial is to see that in doing this he perhaps avoids the problem of the normativity of reasons, only to end up with a problem of moral reasons arbitrariness with respect to settling the telos of a plan. To see the fundamental problem, remember that Gibbard’s psychological description applies to the psychological state of just anyone. It’s not merely that “I count a consideration in favor of doing X (as a matter of my psychological state),” but that the same may be said for anyone at all. All people have

243 Gibbard, *Thinking How to Live*, 189-90; emphasis added.

244 This is reminiscent of Quine’s call for a “naturalized epistemology” where the normativity inherent in epistemology would no longer disturb materialist sensibilities: “Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science” (W.V. Quine, “Epistemology Naturalized,” in *Knowledge: Readings in Contemporary Psychology*, ed. Sven Bernecker and Fred Dretske (Oxford: Oxford University Press, 2000), 273). In line with such an aspiration, in Gibbard’s account a descriptive psychology stands in for the normativity inherent in reasons. One uses a word “reasons,” but in a purely proprietary way as a technical term that indicates the psychological state of proposing to do something.
reasons in this purely descriptive sense: Mao and Mother Theresa both had their plans.\textsuperscript{245} And so, when Gibbard glosses the idea of a reason, $R$, as purely psychological state of a being who plans to do something, emphasizing that the “reasons” tied to such a plan must not be “explained as, somehow, acting as if one believed that $R$ really does weigh in favor of $X$,”\textsuperscript{246} it is important to realize that he has opened the floodgate for “reasons” to spill out all over the place. Simply being in the requisite psychological state of planning to do something is enough to have a “reason” for one’s action. The purely psychological activity of any particular planner calls forth a reason; the reasons do not exist, somehow, independently of people’s different particular plans to do anything at all. \textit{Planning precedes reasons}, for Gibbard. But Gibbard, as a non-cognitivist, rejects full-information/ideal observer theories that might show that the \textit{telos} of a plan is irrational, and his materialism dictates that there can be no \textit{telos} that is somehow given to all planners; any \textit{telos} first comes into being in the endorsement of some particular primate or another whose psychological state in endorsing the course of action may be described (at least in principle, if not in practice) in terms of a materialist Science. And, of course, \textit{any} such \textit{telos} may be so described (the crucial question is whether they can be circumscribed in any meaningful way). Of course Gibbard must say something along these lines if he is to avoid the objection that he has illicitly helped himself to normativity, but it is important to see that in attempting to avoid normativity by replacing reasons with an activity of planning that may be described without remainder in psychological terms, he has fallen headlong into the ditch on the other side of the road: Moral reasons arbitrariness. In this case, Gibbard is seen to be a non-

\textsuperscript{245} In the lyrics of the immortal Alan Jackson, “I got my own reasons why I do what I do,” (Alan Jackson, “She Likes It Too,” \textit{A Lot about Livin’ and a Little ‘Bout Love} (Warner Brothers, 1993).

\textsuperscript{246} Gibbard, \textit{Thinking How to Live}, 189.
objectivist regarding morality. But such objectivity is central to human moral discussion, and if I have successfully cast doubt on norm-expressivism’s ability to capture some significant measure of this objectivity, that is a significant problem for Gibbard’s project.

Gibbard’s treatment of reasons here is consonant with what has already been seen of his thought about reasons in *Wise Choices, Apt Feelings*. There it was pointed out that Gibbard firmly subordinates *reasons* to a person’s *acceptance* of norms. To call R a reason for doing X was just to express one’s acceptance of a system of norms that calls for treating R as favoring X.247 Reasons thus lacked the ability to guide the choice of the system itself; rather, they flowed out of the system of norms upon the acceptance of that system. And this is the same problem Gibbard continues to face in *Thinking How to Live*.

But there is another, perhaps deeper, problem here for Gibbard than just his norm-expressivism. A fundamental difficulty can be seen as Gibbard tries to find what he calls a *deep vindication* for “our” sense of what to do: Do we have any fundamental reasons to trust our planning capacities as “tracking” something important? “The deep puzzle concerns our capacities to ‘track’ the right ultimate ends rather than something else, our capacities to track in our plans, the property that constitutes being okay to do.”248 The worry about “deep vindication” concerns whether we have reason to trust our planning capacity as reliable. This forces a key question: “Reliable with respect to what?” When Gibbard gives a brief Darwinian account of why we have a deep vindication of our capacity for vision, we have a clear sense of what it is tracking and how it could be reliable.249 There can be no such sense

---

247 See notes 91, 92, and 243 above.

248 Gibbard, *Thinking How to Live*, 258.

249 See note 65 above for further development of these questions.
with respect to our planning capacities. Given Gibbard’s Galilean world, what “property” could human planning capacities be reliably tracking if not something having to do with enhancing capacities or chances to reproduce? Furthermore, how could they be tracking this property—whatever it is—when the fundamental moment of settling the *telos* of a plan seems to be the *choice* of a *telos*? This kind of *choosing* does not seem to be responsive to whatever property is putatively being “tracked.” Indeed, it seems that much of the point of this way of thinking about things is to create space between the *telos* of our plans, on the one hand, and our evolutionarily selected nature and the reprehensible things that nature might “track,” on the other.

“Our problem,” Gibbard observes, “is whether we can see why, for beings like us, finding things to be of value should go with their genuinely being of value.”\(^{250}\) What makes this problem so difficult for Gibbard is not simply his expressivism, but his Galilean metaphysic—his materialism. Gibbard is well aware of the problem. At root, materialism gives us a world that itself could only possibly offer a deep vindication for a planning capacity that tracks a property that leads to plans we find ridiculous or (for some *very odd* reason) repulsive. As Gibbard puts it, “If what’s ultimately to be sought in life were one’s genetic reproduction, *that* would explain the correlation [between reproductive success and what’s to be sought in life]—but such a view, I take it, would be ridiculous.”\(^ {251}\) The only planning capacities that might be deeply vindicated by the selection mechanisms that gave rise to us are those that track our reproduction. Our plans might track reproductive success. But to think that our reproduction just is what is to be sought and to plan for it is “to plan

\(^{250}\) Gibbard, *Thinking How to Live*, 264; emphasis in original.

\(^{251}\) Gibbard, *Thinking How to Live*, 264; for the phrase in brackets, see 263.
bizarrely.” But once we have (wisely) ruled this out, what other kind of planning capacity might enjoy deep vindication within a Galilean world? The only thing these capacities might reliably track within such a world has been ruled out as ridiculous or reprehensible. If by some very strange and unexpected twist in the Galilean world “ultimate worth comes apart from maximizing the long-run reproduction of one’s genes,” the only plausible candidate for something that our evolutionarily selected planning capacities might “track” has (mercifully) vanished. In that case, according to Gibbard, there is nothing on the horizon for our planning capacities to track in the same way that our visual capacities track the shapes and movements of objects in the world around us. What, however, will fill the void? The only candidate seems to be the surd of our own pure planning capacity to set a telos for ourselves. But this very capacity has come apart from the only guidance it could have from without. What will guide its application in our plans? Gibbard himself sullenly notes that “no deep vindication will be found for our claims to know how to live—and this, I admit, makes prospects for common inquiry more dicey for living than for scientific inquiry.”

Where does all this leave Gibbard’s quest to do some amount of justice to the “objective pretensions” of normativity? To come right to the point, things do not look good. One of the toughest problems for Gibbard he addresses under the fairly innocent-sounding label of “indexicality.” We have seen how throughout Thinking How to Live Gibbard has relied on hyperdecided idealized planners, like Hera and Zeus. And yet these idealized planners have disagreed about what to aim at in life, and this is all the more the case for you and I, even in the idealized versions of ourselves that Gibbard asks us to

252 Gibbard, Thinking How to Live, 265.

253 Gibbard, Thinking How to Live, 285-6.
imagine. Gibbard states that something is “i-okay… just in case it has whatever property I would now settle on as my guide to living if I thought about the matter in a way that was ideal.”\textsuperscript{254} With this definition in hand, we can start to understand the problem of indexicality. Gibbard asks us to imagine two normal planners in their idealized state, Plato idealized and Xanti idealized, planning about what the thing to do is if one is Socrates faced with the choice to either drink the hemlock or flee. “Xanti idealized, alas, may plan differently from Plato idealized.”\textsuperscript{255} This means that while Plato holds drinking the hemlock to be i-okay, he will understand that for Xanti fleeing will be i-okay. They will never meet at a point where there will be a resolution. “The culprit,” Gibbard points out, “seems to be the indexicality of being i-okay.”\textsuperscript{256} In other words, the “i-okayness” of something depends on idiosyncrasies about the individual that fundamentally shape what they will plan to do—even in an idealized state. And these divergent plans are not susceptible of any resolution; moreover, the plans themselves arise out of factors that—from within the planning perspective itself—are contingent or haphazard. Since picking out a telos for one’s plan is not beholden to reasons at all, and a fortiori not to reasons addressed to all planners equally, the idiosyncrasies of one’s birth and upbringing will not be susceptible to correction by reasons. And one’s change (not correction) in plan in response to the “reasons” addressed to oneself by others owes in no small measure to the contingencies in one’s planning that cause one to give weight to this and not to that as a “reason.”

\textsuperscript{254} Gibbard, \textit{Thinking How to Live}, 240.

\textsuperscript{255} Gibbard, \textit{Thinking How to Live}, 243.

\textsuperscript{256} Gibbard, \textit{Thinking How to Live}, 243.
If Gibbard is right that “some such indexicality is unavoidable,” and if he is right that “i-okayness” will vary from person to person due to indexicality, then there are serious problems for the objective pretensions of normativity. Remember the three pretenses Gibbard wanted to defend in *Wise Choices, Apt Feelings*: (1) the independence of a norm from one’s own acceptance of it, (2) the idea that one’s acceptance of a norm is for universally recognizable, non-arbitrary reasons rather than particularities of one’s nature and upbringing, and (3) the interpersonal authority of normativity.

Let’s start with some of the ways Gibbard’s norm-expressivism gives an unsatisfactory treatment of the pretense of non-arbitrary reasons. For starters, the deep influence of indexicality just is the negation of the second pretense, for the problem of indexicality is saying that even idealized versions of us will inevitably be grasped by the idiosyncrasies of their situations. Even idealized planners accept things not for reasons recognizable by anyone, but because they are in the grip of various contingencies. To make the problem here all the more plain, think back to the idealized anorexic we met in *Wise Choices, Apt Feelings* and the idealized criminal that Gilbert Harman placed in service of his moral relativism. The “ideally coherent anorexic” holds that starving herself for a thin figure is i-okay, while the idealized criminal holds that murdering Ortcutt is i-okay. The telos of their respective plans points them in these directions, and even their idealized selves might well hold these same plans of action to be okay.\(^\text{258}\) Indeed, how could the idealized anorexic’s

\(^{257}\) Gibbard, *Thinking How to Live*, 244.

\(^{258}\) That Gibbard holds out little hope for substantive constraint from content-neutral “ideal rationality” is argued extensively above. Again, “sheer coherence places few constraints on who can be left out and why; the constraints with teeth are the ones of cost,” (Gibbard, *Wise Choices, Apt Feelings*, 219). Recall, also, Gibbard’s thoughts on Caligula. For Gibbard it is not full information/ideal coherence that rules out Caligula, but content-specific material that finally finds its ground in the pragmatics tied to our evolutionary psychology. This, I think, is also borne out in *Thinking How to Live* (though there is much less emphasis on evolutionary
telos be corrected? This is the problem of indexicality that Gibbard identifies. The severe difficulties Gibbard encountered in trying to work out a plausible account here have already been seen above, and nothing in *Thinking How to Live* cases the difficulty for him.

Quite the opposite. Thinking about the indexicality problem, Gibbard states that “all this means that in an important sense, Plato thinks that he and Xanti can’t form a community of inquiry on questions of what to do if Socrates.” The situation is nothing like the situation we find in scientific inquiry, where a community of inquiry forms around a uniform “outward clash” with a stubborn and constant reality. Instead, Gibbard identifies two kinds of impasse that end the possibility of normative discussion about the “thing to do.” A constitutional impasse occurs when someone takes the other person not to be lacking in any epistemic virtue, but rather to be constitutionally unable to “see” the thing to do. A multi-equilibrium impasse occurs when two parties are ideally coherent with respect to what their normative positions are (in Rawls’ terms, they are in “reflective equilibrium”), but each finds the other’s equilibrium disagreeable. Faced with these two impasses, Gibbard goes looking for partners in the crime. But there is a crucial difference. Cognitivists of various psychology in the later work), where—as seen above—Hera and Zeus and idealized Plato and idealized Xanthippe are fully-informed and ideally-coherent, and yet disagree on what is the thing to do. See also the problem of a “constitutional impasse” and a “multi-equilibrium impasse” below.

---


260 Gibbard’s talk about communities of inquiry seems to be influenced by Charles Sander Peirce, who held that the uniformity of the “no” that would cut off theorizing in science would call out a community of inquiry that gradually found agreement on scientific matters due to this uniform pressure. I have previously addressed Jürgen Habermas’ attempt to devise a moral theory from the idea of communities of inquiry by devising a substitute source of the uniform “no” central to scientific communities on Peirce’s view. See Seeman, “Peirce, Habermas, and Moral Absolutes.”

261 Gibbard, *Thinking How to Live*, 269.

stripes hold that there are reasons available for deciding these matters. Additionally, there are theists who hold that there are reasons addressed to all people by a God who stands over against all people equally. Of course all those who hold such positions may be wrong, but their problem is not at this point. Such positions have clear answers to the problem Gibbard faces here.

Gibbard clearly lacks a satisfactory quasi-objectivity with respect to the pretense of non-arbitrary reasons and the pretense of independence fares no better under his ministrations. His answer here has not changed from Wise Choices, Apt Feelings. “The claim of independence, then, turns out to be internal to normative thinking—though arrayed in sumptuous rhetoric.”

263 We cannot avoid taking our own plans as being worth following; this is part of what it is to plan to do something. The problem here is that we see that other people have identical reasons to take their own plans as worth following—namely, that they are their own plans. 264 This is not a very interesting reason, and not what people (or at least not what I, and I believe most others) mean when they think that the norms they accept are independent of their act of accepting them. I, at any rate, mean to be saying something much more robust when I say that sex trafficking of girls is wrong. Moreover, it is clear that both (1) Gibbard’s expressivism with respect to settling the telos of a plan and (2) the impossibility of understanding one’s own appeals to reasons as anything more than “sumptuous rhetoric” adorning one’s preferences when there is an impasse with respect to

263 Gibbard, Thinking How to Live, 186.

264 Here, again, a full-information/ideal-rationality descriptivism might offer a way out of this problem, though Gibbard sees other problems with moving in such a direction. The criticism of different plans in terms of full-information and ideal-coherence is limited in Gibbard’s non-cognitivist norm-expressivism. In any case, Gibbard allows little, if any, room for rational criticism between Hera and Zeus or between ideal Plato and ideal Xanthippe. Ex hypothesi, they are fully informed and ideally consistent. As seen above, the same holds for the “ideally coherent anorexic” of Wise Choices, Apt Feelings.
one aim versus another, that both of these directly undercut the independence of normative judgment Gibbard wants to salvage.

Finally, the pretense of interpersonal authority is also clearly undercut by Gibbard’s project. Given the idiosyncrasies at the heart of our planning activity and the impossibility of appealing to independent reasons to elevate the telos of one plan over the telos of another, Gibbard has no clear way to establish the authority behind people’s plans—either to people themselves or to others. The unadorned preference for one telos is pitted against another such preference.

In sum, the worry here is fairly easy to grasp. What about conflicting plans and crazy plans? When Zeus and Hera’s plans conflict or when Ares’ plan aims at a telos that is crazy, Gibbard’s expressivism very quickly leads to a bald (“pure planning”) human activity that does not respond to reasons, but rather constitutes “reasons” through human activities responsive to non-normative forces. The resulting arbitrariness is installed at the very heart of the plans at the center of Gibbard’s project—namely, in the telos at which those plans aim. There is a consequent failure to do anything satisfactory with the “objective pretensions” that Gibbard rightly understands as central to human normativity. In the end, Gibbard strips moral reasons of their independent, authoritative prescriptivity and leaves us with “reasons” tied to the system of norms a person endorses or a plan she makes. Reasons flow out of what a person endorses or plans to do and so cannot constrain or inform the activity of endorsement or planning. This is moral reasons arbitrariness.

Behind the views that lead Gibbard into this coal pit, it is not difficult to see materialism exercising a vast influence on his thought. In keeping with his Galilean metaphysic, Gibbard holds that “in important realms of science, a deep vindication of our
powers of correct judgment seems to be in the offing, whereas there is no such plausible prospect for questions of how to live.”

Since the only deep vindication is one that should be repulsive to us precisely in our capacity as planners, we are cast back onto the thin resources of our own activity operating in a reasons vacuum. Reasons have gone missing, leaving Gibbard with a problem of moral reasons arbitrariness infecting his norm expressivism.

In the final analysis, moral reasons are undermined in Gibbard’s account. As Gibbard sees matters, for those who believe in a Galilean metaphysic normativity must somehow arise from human beings themselves, since the basic stuff (quarks, quantum packets, strings or the like) is clearly not such as to address normative matters at all and (as we saw Gibbard arguing above) the Galilean world does not have room for normative or rational-making facts. If Gibbard is right (and a good number of philosophers—full-information descriptivists, for example—think he isn’t), the Galilean materialist must look to the activity of human beings for the constitutive ground of normativity. I will not argue here that Gibbard is right about this, but if he were, this would seem to limit the resources for moral philosophy to one of two broad, diversely conceived positions: either (1) something about how the evolutionary processes that gave rise to human beings (or other contingent rational beings) imposed brute constraints on human behavior that gave rise to the “objective pretensions” of normativity (including morality) that explain normativity and in any event serve all the purposes needed to sustain our cooperative endeavors, or else (2) something about humans (and any other contingent rational beings) allows them to constrain themselves as beings who have (somehow) become rational, where that just means

265 Gibbard, Thinking How to Live, 285.
that we are able to reflectively consider our own behavior, and that this reflective standpoint underwrites certain procedures that ground human normativity simply as the view from the reflective stance that constitutes our behavior. This second way of proceeding involves turning inward to sui generis human moral willing; this is the tack that Korsgaard (and many other theorists) pursues, and it will be examined in Chapter 4. Gibbard, on the other hand, pursues the first tack of turning outward to the forces of the material world itself, and the present chapter has shown the insurmountable difficulties Gibbard encountered in trying to give an account of normativity on this basis. Gibbard has an abundance of external causal forces that cut across the subjective motivational sets of all people (and indeed give rise to those SMS’s), but precisely what goes missing in all of this is the independence and authority of normativity itself. Gibbard’s materialist norm-expressivism fails to retain any meaningful moral objectivity, despite his best efforts to salvage something interesting from the three “objective pretensions” of morality, leaving the materialist norm-expressivist with “moral reasons” that are neither independent nor authoritative nor non-arbitrary. Causes there may be in abundance—but causes are not reasons, and these causes clearly fall out on different sides of the various moral issues that press on us.

If Gibbard’s Galilean account says all there is to be said for morality, then there are no independent, authoritatively prescriptive moral reasons as such and his moral philosophy fails to sustain morality in its own terms. Gibbard, in short, captures a kind of externality located in Galilean causal forces that cut across all people’s SMS’s, but external moral reasons

266 Something of the disagreement between Gibbard and Korsgaard in these matters can be seen in the article referred to in note 50 above. This disagreement will be explored more fully in Chapter 4.

267 As was made clear in the Introduction, no claim is being made here that these two tacks—inward and outward—are exhaustive. Perhaps there are other ways forward that are not versions or hybrids of these two. It is not the burden of this dissertation to argue this point one way or the other.
drop out of his account, being submerged in Galilean causes without remainder. And so, in his efforts to pursue one possible tack open to the Galilean materialist, Gibbard succumbs to moral reasons arbitrariness. Gibbard might not take this to be a fatal objection to his moral philosophy, but if I have successfully cast doubt on the success of Gibbard’s assiduous efforts to “save what is clear” about rationality as it is applied to normativity and to salvage a significant moral objectivity—that is a significant blow to Gibbard’s norm-expressivism and other relevantly similar moral philosophies. After considering Korsgaard’s attempts to do better by following a second tack, it will be argued that Gibbard provides an important clue for any attempt to avoid moral reasons arbitrariness: it’s not enough to have externality that cuts across all people’s SMS’s—that externality must address moral reasons to moral agents. Externality without reasons still leads to moral reasons arbitrariness. I have argued that Gibbard is a clear example of this.

III. External Reasons and the Second Person Address in Gibbard and Darwall

In this concluding section of Chapter 3, the focus will stay on reasons. I have tried to show that Gibbard’s moral philosophy lacks any robust account of reasons. Reasons tag along behind the crucial human activity of endorsement/planning that does the heavy lifting in Gibbard’s account of human normativity, and thus they end up with nothing to do but express acceptance of a system of norms or some plan or another. The crucial moments of endorsement or settling the telos of a plan are isolated from the force of reasons. As a result, Gibbard’s norm expressivism fails to give substance to the “objective pretensions” of normativity and leads into a coal pit of normative reasons arbitrariness.
A more hopeful alternative to Gibbard’s take on reasons emerges from a comparison of Gibbard’s view with that of Stephen Darwall. Darwall’s account focuses on the notion of “second-personal address” and is open to a robust externalist understanding of reasons. At the heart of normative reasons stands the summons of the other person to respond to the reasons she offers up for my deliberation. While it will be argued at the end of Chapter 4 that finally Darwall betrays his own best insights, here the focus will be on the sharp contrast between his views and Gibbard’s views. Interestingly, it turns out that Gibbard dubs himself a “reasons externalist” (in a sense to be specified below) and reflects on the second personal aspect of reasons. Let’s revisit Gibbard’s views on reasons one last time before considering Darwall’s alternative.

One might be surprised upon reading Gibbard’s claim to be a reasons externalist in an article on Bernard Williams’ well-known critique of reasons externalism. Both Wise Choices, Apt Feelings and Thinking How to Live lend themselves much more readily to reasons internalism, at least on first blush. It turns out, however, that reasons externalism is a pretty big tent on Gibbard’s understanding of the idea. When Gibbard calls himself a reasons externalist, he does not mean that there are reasons that are somehow independent of us that one can either see or fail to see (and so fail to be motivated by them). If talk of “external reasons” is tangled up in this sort of “mysterious Platonism,” says Gibbard, then it would rightly “be laughed aside.” Gibbard’s rhetoric fits the (admittedly odd) views of G.E. Moore, which do seem to be a rather odd Platonism appended to a Materialist metaphysic. Gibbard’s language echoes that of Mackie’s critique of “queer” things in the universe.

---

268 Gibbard, “Reasons Thin and Thick,” 293.

269 Gibbard, “Reasons Thin and Thick,” 294. Gibbard’s rhetoric fits the (admittedly odd) views of G.E. Moore, which do seem to be a rather odd Platonism appended to a Materialist metaphysic. Gibbard’s language echoes that of Mackie’s critique of “queer” things in the universe.
trying to figure out how they might apply more broadly than just to the individual and his or her own subjective motivational set, we cannot turn to a laughable “Platonism” for help.

Nor does Gibbard think that we can turn to a descriptivist sort of naturalism. Gibbard holds that there is an imperatival element in our giving of reasons that cannot be captured in this way. There is, as Gibbard sees things, no “naturalistic synonym” for this imperatival element of morality. Instead, as has already been seen with Gibbard, we must be the source of a non-natural concept that arises from our activity of thinking what to do. I plan, and a concept is thereby spread over all that falls within the ambit of my plan. Thus, to use Gibbard’s example of a man who is motivated to kick me coming to see that there is a reason for him not to kick me, this man’s coming “to believe that my pain is reason not to kick me would be for him to accept a directive to weigh my pain in favor of desisting.” Once again, we see reasons flowing out of our activity of accepting a directive, a system of norms, a plan: endorsement then reasons, this is Gibbard’s pattern.

All of this fits well with what we have already seen of Gibbard’s thought. The only puzzle here is how Gibbard can call such reasons external. After all, it has been shown above how this moment of acceptance, bereft of all response to reasons, is highly exposed to the influence of the contents of our “subjective motivational set” (though that terminology was not employed), which is Williams’ classic language for internal reasons. The “reasons” flowing out of such a situation would seem to be internal reasons par excellence.

As seen above, Gibbard’s argument is that “the special element that makes normative thought normative… involves a kind of endorsement—an endorsement that any descriptivist analysis treats inadequately,” (Gibbard, Wise Choices, Apt Feelings, 33). According to Gibbard, this “single loophole remains unpluggable by descriptivist analysis,” (Gibbard, Wise Choices, Apt Feelings, 33). Gibbard’s lineage, tracing back to R.M. Hare, is clearly displayed here.


Gibbard, “Reasons Thin and Thick,” 295; emphasis added.
Strangely enough, the key to understanding Gibbard’s putative “externalism” here is to think of the matter strictly internally and to take judgment internalism at full strength. My very acceptance of a system of norms and the reasons embedded within it necessitates my viewing it as holding, as being rightly applied; this is part of what it is to accept it. I myself, in my having accepted the system of norms and its reasons, just have taken there to be a reason, and part of what this means is that (I take it that) there is a reason for the man to stop kicking me, even if he does not accept it. Thus Gibbard states that “thoughts of ‘considering matters aright’ are a part of self-conscious planning. I can say, ‘To consider matters that way is to consider them aright,’ and that roughly amounts to the imperative, ‘Consider matters that way!’ If we cannot renounce such imperatives, we cannot deny all sense to talk of ‘considering matters aright.’” We accept a system of norms and say, “you do so as well.” And, voila: an “external” reason.

Gibbard’s hyper-internalism at this point is what makes sense of what would otherwise seem strange. In a key paragraph we find him speaking of sincerity twice, which would typically be a strange think to talk about in defending a notion of external reasons. As our man thinks about kicking us, we address reasons to him: “You should weigh my pain against kicking me,” we say to him. Gibbard notes, however: “Vain hope that he will accept this, but I am being sincere: I myself accept it.” This is not something a reasons externalist would normally find relevant to a defense of external reasons, but for Gibbard this is key. Your sincerity in accepting it and in offering it up is just what renders the reason external, in

---

273 This is closely tied to Gibbard’s treatment of the pretense of independence.


275 Gibbard, “Reasons Thin and Thick,” 291; emphasis added.
the sense that you have to offer it to the man who wants to kick you as a reason. You mean it, after all.

There is, of course, a problem here. More than one can play this game, and kicking man can help himself to the same way of proceeding. His reasons too are external for him. He, too, is sincere in offering his glee in venting his rage on someone who can’t resist as a reason for his behavior. “Vain hope,” he might say, “that this guy will accept it. But I am sincere: I myself accept it.” And voila: another external reason! One merely needs to shift the viewpoint around.

The upshot is that Gibbard’s “reasons externalism” ends up looking a lot like reasons internalism with a little twist about how we have to view our own “reasons” that flow from the system of norms (or the plan) we accept. In fact, it is normative reasons internalism. Until Gibbard finds a way to make our activity of acceptance itself responsive to reasons rather than being what first gives rise to our reasons, he will not be able to give a satisfactory account of the role of reasons in human normativity. And, as a result, he will fail to give even a remotely satisfactory account of the “objective pretensions” of normativity. Gibbard’s normative reasons internalism leads straight to normative reasons arbitrariness—including moral reasons arbitrariness.

By way of contrast, Darwall offers a view in which reasons are rooted in second-personal address. The idea is that you have the authority to make a claim on me, and that your addressing a reason to me renders me accountable. “Our dignity as persons includes… an irreducibly second-personal authority to demand respect for this very authority and for
the requirements with which it gives us the standing to demand compliance.” Darwall’s best insight here is that there is something about the dignity of the other that does in fact give me a reason to act when he makes a claim on me. The person gives the reason and it takes its authority as a reason from the dignity of the person giving it. Moreover, the reason can be genuinely external to me, and not in any bad “mysterious Platonism” way; for it is a person who is the source of the reason, not a “form” or a queer, indefinable “non-natural property” we somehow intuit. The person who addresses a reason to me—who summons me—has the standing to make this claim, even should I fail to recognize the claim. And the reason carried by his address is there, whether I recognize it or not. Here, unlike Gibbard’s account, we have the possibility of genuinely external reasons—reasons that stand independently of me and my motivations.

It should also be emphasized that not only are second-personal reasons external, they are also genuinely reasons. Throughout his book Darwall disparages views where the “reasons” on offer are “reasons of the wrong kind.” “To be a reason of the right kind, a consideration must justify the relevant attitude in its own terms. It must be a fact about or

---


277 Darwall may well, however, face difficulties grounding this assumption within his materialist Kantianism.

278 Darwall briefly considers Bernard Williams’ argument for internal reasons, and argues that his own proposal could be made consistent with “internal reasons” if the idea of a subjective motivational set were carefully qualified. If one were to take this route, Darwall insists (very much in line with Kant’s insistence that morality must avoid “pathology”) that whatever subjective motives we have should not be determined by objects (this would render them heteronomous), but only by principles. See Darwall, *The Second-Person Standpoint*, 94-5. This is part and parcel of the Kantianism that ends up vitiating Darwall’s own best insight. This issue will be treated more fully in Chapter 4.

279 In all of this I will not concern myself with arguments against the very idea of persons such as those advanced by Derek Parfit. See Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984). For a brief summary and reply, see Thomas Nagel, *The View from Nowhere* (Oxford: Oxford University Press, 1986), 43-5.
feature of some object, appropriate consideration of which could provide someone’s reason for a warranted attitude of that kind toward it.\textsuperscript{280} Thus, to use an example Darwall himself employs, the fact that it would be desirable to hold a belief (say, that your senior colleagues are acting wisely), that desirability itself is the wrong sort of reason for holding the belief. It gives no warrant to it; the terms do not line up, so that the one isn’t even speaking to the same issue as the other.

Darwall rightly sees “the wrong kind of reason” problem as a serious difficulty for moral philosophy. What can give us—precisely as moral persons—moral reasons for acting? Where will we find the right kind of reasons to speak to us qua moral agents? What sorts of reasons can be couched in the right terms to address us in that standing? Darwall answers: Only a person can do this. An oft-repeated slogan in The Second-Person Standpoint runs, “Second-personal authority out, second-personal authority in.”\textsuperscript{281} The reasons we can respond to as moral persons must come from a moral person—either oneself or another. Personhood has to be introduced to give a reason of the right kind.

If this is true, it explains a lot of the troubles Gibbard bogs down in. Having a reason of the right kind seems to be exactly Gibbard’s problem. Having started out with materialism and taken that starting point seriously, he cannot come up with the right kind of reasons. Causes he has aplenty, but this is not what a moral agent needs. If Darwall is right (as I believe he is), those lead only to reasons of the wrong kind, and the key moment in Gibbard’s materialist expressivism is a human activity that ends up isolated from the authority of reasons and thus subject to all sorts of adventitious forces. The human activity

\textsuperscript{280} Darwall, The Second-Person Standpoint, 16; emphasis in original.

\textsuperscript{281} Darwall, The Second-Person Standpoint, 59.
giving rise to “normativity” ends up beholden to forces that have nothing to do with normativity itself. The difficulty thus facing Gibbard the Galilean materialist, is how to transmute the non-personal Galilean forces that explain persons into something like normative reasons—quasi-reasons. How will the Galilean naturalist get second-personal authority out when third-person explanation goes in? Gibbard himself notes the problem: “If we try to paint normative life as a part of nature, crucial parts keep looking off shape. Reasons in the picture look not quite like genuine reasons…”

Like Gibbard, Darwall is a metaphysical materialist, and it will be argued below that he too finally fails to transmute the material stuff into second-personal reasons. But the purpose of this concluding section of Chapter 3 has not been to critique Darwall’s project, but to announce his best insight and to suggest that it may offer a way forward. No doubt many possible objections to second-personal authority and dignity as the grounds for external moral reasons have sprung up in your mind. There are many such objections, and some of them are quite damning—chief among them being the problem of the lack of uniformity in the claims addressed to us by others and how they are to be smoothed out into a coherent moral address rather than a cacophony of discordant voices. Darwall’s answer will be a fundamentally Kantian answer that will end up bringing his project much closer to that of his fellow Kantian, Christine Korsgaard, than he would like. In both cases it turns out that first-personal authority swallows up second-personal authority, resulting in severe problems of how first-personal authority is to be limited in a principled way.

But again, the purpose here is not to critique or to defend Darwall at this juncture, but rather to announce a different way forward and to remind us of what has been seen to

---

282 See the epigraph for the dissertation.
this point. Gibbard’s failure to do justice to the “objective pretensions” of normativity and avoid moral reasons arbitrariness owes to his inadequate account of moral reasons and their authority in providing normative guidance. A possible analysis of a more general problem of which Gibbard’s difficulties are merely a species is provided by Darwall: In order to avoid normative reasons arbitrariness we need moral reasons that are both external and of the right kind—where being of the “right kind” has to do with the address of a person who has the dignity and standing to give us reasons.

It is a mark of Kantianism to be very concerned that we have the right kind of reasons for our moral actions. Otherwise our actions end up not being moral at all. I share this concern and thus have more sympathy for the Kantian tack than for a project like Gibbard’s. One of the most prominent Kantian theorists today is Christine Korsgaard, and it is to her project that we will now turn. Korsgaard’s project is unflinchingly first-personal, and it will be argued that she is right that a Kantian with her metaphysical persuasions is committed to this and that within such an outlook Darwall’s second-personal project finally collapses back into the first-person standpoint. The main difficulty will be how to limit choice within that first-person standpoint where someone addresses reasons to herself. Interestingly, as a “Galilean” materialist, Korsgaard—like Gibbard—puts a human activity at the heart of her project: “Value springs from the act of rational choice.”

It turns out that endorsement is no less central for Korsgaard than it is for Gibbard, but Korsgaard will emphasize that what matters is reflective endorsement. It will be crucial to examine just what reflection is able to add to endorsement, and if it will suffice to avoid the moral reasons arbitrariness Gibbard plunges into.

\[^{283}\text{Korsgaard, Creating the Kingdom of Ends, x.}\]
In her introduction to Christine Korsgaard’s thoughtful and challenging book, *The Sources of Normativity*, Onora O’Neill states that Korsgaard “attempts to locate normativity within rather than beyond human activity.”¹ This should sound familiar from our consideration of Allan Gibbard’s norm expressivism, a view taking as its key doctrine the idea that a distinction between belief and the activity of planning is “the contrast that lies at the heart of expressivism.”² It is not surprising that Gibbard lauds Korsgaard’s decision to ground her moral philosophy in the insight that, as she puts it, “moral properties are projections of human dispositions.”³ Though Gibbard and Korsgaard differ regarding how such projections are constrained, Gibbard believes they share the same starting points.⁴ Such expressivist “starting points will be controversial,” Gibbard notes, “but not with me; I applaud.”⁵ Indeed, Gibbard goes so far as to claim Korsgaard for the expressivist camp, although he may be overreaching here. Yet he does have a point; for the human activity of

---

² Gibbard, *Thinking How to Live*, 76.
⁴ The differences between their accounts will emerge very clearly below, including why Korsgaard thinks that human expressions of morality are not only expressions.
⁵ Gibbard, “Morality as Consistency in Living,” 141.
endorsement is every bit as central for Korsgaard as it is for Gibbard—in some ways even more so.

The divide between Gibbard and Korsgaard centers on the way the human activity of endorsement is constrained or limited so that not just anything will be endorsed. While for Gibbard the brute constraints of our evolutionary history play a central role in what we endorse in our planning activities, for Korsgaard it is reflection that carries the load. Following what she takes to be Immanuel Kant’s position, Korsgaard asserts that “rational choice itself makes its object good. [Kant’s] idea is that rational choice has what I will call a value-conferring status.” It is unclear how such reflection is supposed to be related to what Korsgaard calls the “Modern Scientific World View” and the efficient causes central to its account of human beings. As she wrestles with these problems, Korsgaard ends up struggling with the same type of problems Kant wrestled with in his Critique of Judgment, problems neither philosopher successfully resolves. Furthermore, even if such problems could be resolved, it is not clear that Korsgaard’s strategy for limiting the human activity of endorsement succeeds any better than Gibbard’s strategy.

The first section of this chapter will outline the main ideas in Korsgaard’s moral philosophy, drawing out similarities and contrasts with Gibbard’s account as appropriate. This expository task is not easy. Gibbard, an acute and subtle philosopher in his own right, confesses that the argument Korsgaard presents in The Sources of Normativity “is rich and difficult, and though I’ll attempt a reading, I very much can’t feel assured that I’m getting it right.” One can only agree with Gibbard’s assessment of the challenges Korsgaard’s work

---

6 Korsgaard, Creating the Kingdom of Ends, 122.

7 Gibbard, “Morality as Consistency in Living,” 141.
presents. In offering my own reading of Korsgaard’s work in the first section of this chapter, I will take Korsgaard’s *The Sources of Normativity* as the main source for understanding her moral philosophy, while also drawing on Korsgaard’s many important articles and book chapters to fill in the picture. A particular concern of the first section is to explore the way Korsgaard understands reasons, and to examine the relationship between reasons and the “practical identities” central to her thought in *The Sources of Normativity*. These practical identities mediate both (a) the activity whereby we “confer” value on elements of the world, and (b) the reflection that limits what can be endorsed by human beings as *human beings*. The first section concludes by considering the central role the notion of a maxim plays in Korsgaard’s thought, arguing that she presents two conflicting notions of maxims. For the purposes of this chapter these differing understandings shall be called identity-priority maxims and form-priority maxims. Roughly, if a maxim is the proposal that has a willing “*to-do-this-act-for-the-sake-of-this-end*’ structure,” identity-priority maxims are maxims where the endorsement by a rational agent in what Korsgaard calls the “test of endorsement” confers value on the act-end pair. Form-priority maxims, on the other hand, are roughly maxims where the form of the maxim itself renders any endorsement taking that form rational, thereby conferring value on the act-end pair. These rough summaries of Korsgaard’s two different notions of the maxim and its role in moral philosophy will be unpacked below, with much attention to Korsgaard’s own formulations. The point here is merely to sketch the direction that will be pursued below. Korsgaard will be seen in

---

subsequent sections to face different problems depending on which notion of a maxim she employs. Either path leads to moral reasons arbitrariness.

The second section opens by posing a problem for Korsgaard’s moral philosophy centering on her two different ways of conceiving of how value might be conferred on an end through reflective endorsement: the *identity/reasons problem*. On the one hand, if Korsgaard prioritizes identity over reasons, there are no reasons to constrain the identities that an agent takes up—these reasons first needing to have *some* identity or another in place in order to exist in the first place. It would seem that one identity may be taken up with as much reason and as any other. The upstanding, moral citizen will have his reasons, and the contented criminal will have hers, each flowing out of their practical identities. If Korsgaard prioritizes identities over reasons, she faces a difficult struggle of showing how the practical identities in view can themselves be taken up *for reasons*. There is no clear way the adoption of one identity over another will itself be responsive to reasons, since the identities precede reasons. Moral reasons arbitrariness clearly threatens. On the other hand, if Korsgaard prioritizes reasons over identities by emphasizing the form of the maxim, it is not clear that the mere form of a maxim—that an act will be done for the sake of an end—will be able to constrain the will unless that will is already rational.⁹ Here Gibbard is right: there needs to be a substantive “why-stopper.” If the rational identity were somehow already in place, then perhaps maxims would take on the needed form as they were willed by a rational being. But the form of the maxim itself is supposed to be providing the needed constraint on the identity if reasons are to precede identity and thus to constrain identities. Nor is it clear how a *materialist* like Korsgaard would give an account of how the form of a maxim might

---

⁹ These points will be developed extensively below.
preexist or somehow call out the endorsing agent such that by following this one form the agent is *thereby* rational.\textsuperscript{10} There is a typical and highly problematic Kantian circularity to Korsgaard’s moral philosophy. The identity/reasons problem is a very difficult problem Korsgaard must surmount.

The remainder of section two considers the two ways open to Korsgaard to deal with the identity/reasons problem, and critiques Korsgaard’s handling of moral reasons, focusing particularly on the relationship between moral reasons and practical identities. On the one hand, if Korsgaard opts for identity-priority maxims, she faces the problem of correcting or somehow constraining practical identities, when reasons are themselves tied to these identities. As Korsgaard puts it, our practical identities “give rise to reasons and obligations. Your reasons express your identity.”\textsuperscript{11} What, then, will constrain the identities that give us reasons? Are the reasons emerging from one identity just as good as those emerging from any other? Korsgaard attempts to solve these difficulties by appealing to an overarching identity that is performatively necessary for all of us: the reflective identity of a human being that continually pushes us into a regress toward the unconditioned ground of *any* practical identity whatsoever, which turns out to be nothing other than the reflective self. One’s humanity must be valued insofar as we take up any identity whatsoever and thus

\textsuperscript{10} Note 17 in the Introduction gives my reasons for using the term “materialist” instead of “naturalist.” In close proximity to quotations where Korsgaard uses the word “naturalism” I sometimes relax this practice.

\textsuperscript{11} Korsgaard, *The Sources of Normativity*, 101. Korsgaard’s plight here is clearly related to Gibbard’s problem with finding the telos of a plan, a relation that owes to the close affinity of their moral philosophies. For both Korsgaard and Gibbard, we “gild and stain” a world devoid of normativity with values that exist only because of our endorsement.
have reasons to act. Section two first argues that this move fails to resolve the problems that attend Korsgaard’s attempt to identify human practical identities as the sources of normativity.

On the other hand, Korsgaard could opt for form-priority maxims, where the form of the maxim’s structure itself precedes and calls out the correlated identity of “humanity.” Although this is the minority report from her writings, it seems to be gaining momentum in her latest essays. Indeed, Kosgaard’s latest book dramatically confirms this direction in Korsgaard’s thought. Yet this tack also fails to avoid moral reasons arbitrariness. The central problem is that maxims need a settled, non-arbitrary rational identity in the one endorsing them if they are not to be arbitrary. As will be argued below, the mere structure of a maxim as an act done for the sake of an end (even if it could be conceived as somehow itself first drawing out the rational human agent) is not sufficiently thick to constrain the agents who will in accordance with this form. Contrary to Korsgaard, we cannot “tell whether our maxims should be laws by attending not to their matter but to their form.”

Nor will it do to declare by metaphysical fiat that the needed human identity and the

---

12 Following what she understands to be Kant’s position, Korsgaard has a special, technical sense of “humanity” as rational agency and the ability to consciously take up an act for the sake of some end. Her account is sketched below, on pp. 270-73. When “humanity” is brought up in connection with Korsgaard’s moral philosophy below, it should be understood in this way.

13 Christine M. Korsgaard, Self-Constitution: Agency Identity, and Integrity (Oxford: Oxford University Press, 2009). This book was published as final revisions of this dissertation were under way, and I only obtained a copy in July 2009. I have worked in a few references, particularly regarding what this book indicates about the trajectory of Korsgaard’s thought. In general, the book is a remarkable confirmation of the direction I had argued Korsgaard might have been heading. So much so, that I might be willing to say that what I call the “form-priority maxims” view is now her official position.

lawfulness of the maxim structure arise primordially together.¹⁵ Such a move incurs many of the same weaknesses of attempts to prioritize either the identity or the form. Thus the second section concludes that Korsgaard fails to avoid moral reasons arbitrariness.

The third section is an ambitious attempt to unearth a deeper, principled reason for Korsgaard’s difficulties with the reasons/identity problem by exploring the ways in which a classic Kantian problem—the “transition” problem—runs through Korsgaard’s moral philosophy, showing up particularly in her doctrine of “the two standpoints.”¹⁶ While it is not essential to the most basic point of this chapter to defend the connection that I try to draw out between Korsgaard’s attempt to solve her transition problem and Kant’s own efforts to defuse the problem with his notion of “heautonomy” (that is, a certain self-reflexivity of judgment applied to itself and in service of a kind of objectivity), if the connection is successfully displayed, it provides an important context for understanding the difficulties Korsgaard faces and might help make plain a difficulty facing Kantian moral philosophies more generally. If the connection is successfully expliciated, it may point to a typical Kantian reliance on a self-reflexivity that creates more problems than it solves. If successful, the argument of this section would show that Korsgaard’s moral philosophy suffers from her own version of an architectonic circularity that beset Kant and perhaps also his heirs. In Korsgaard’s moral philosophy this circularity takes the shape of the

¹⁵ Such a move would be not unlike a theme in Heidegger’s writings. Though such a stance will not be explored in the text of this dissertation, a number of footnotes will address Heideggerian themes.

¹⁶ The reader will notice that there is a considerable amount of hedging in this paragraph. This owes to the fact that this dissertation is not meant to give anything like authoritative Kant exegesis. Those thickets are just too dense for me to wade into here in anything like a comprehensive manner. Thus, this section of the dissertation is meant as suggestive; and, I want to make clear, its suggestions are not finally necessary for the essential point of the chapter, which is really made in section two. That said, I believe the connections I draw out in this section are at least in the neighborhood of being correct and that they do add significant support to the overarching point of the chapter.
identity/reasons problem. Korsgaard’s two conflicting views of the maxim each prioritize one moment of the circle in an attempt to ground the whole in such a way as to avoid moral reasons arbitrariness.

The final section compares Korsgaard’s moral philosophy with that of another Kantian, Stephen Darwall. After noting some strengths both views share regarding the connection between reasons and persons, this section argues that Darwall’s moral philosophy—although more explicit in developing importance of the second-person address for moral reasons—finally holds that the second-personal address has no authority apart from the reasons that the autonomous agent gives himself in the first-person standpoint. Only as second-personal reasons are ratified within the first-person standpoint do they have any authority. Finally, for all Darwall’s talk about “second-personal authority,” such “authority” is swallowed up by the genuine authority of the first person standpoint of the autonomous agent, in a kind of Kantian fashion (though what Kant himself believed is not so easily pinned down). Darwall’s Kantian proclivities eventually end up overwhelming his insights into the importance of the second-person address for moral reasons as the second-person standpoint collapses into the first-person standpoint. Nonetheless, both Darwall and Korsgaard have insights that point us beyond their own philosophies toward the kind of reasons we need if moral reasons arbitrariness is to be avoided.
I. Acting for Our Reasons: Human Reflection and the Limits of Endorsement

A. Value, Humanity, and the “Modern Scientific World View”

Christine Korsgaard shares with Allan Gibbard a commitment to what Gibbard called the “Galilean Core” of scientific knowledge and Korsgaard calls the “Modern Scientific World View.” She affirms that her moral philosophy “grounds normativity in certain natural—that is, psychological and biological—facts”; she, too, values her “naturalist” credentials. Unlike Gibbard, however, Korsgaard is palpably uneasy about the implications of their common metaphysical commitment for morality, and she is much more resistant to letting that metaphysic dictate the way human morality is to be understood. “The Scientific World View is no substitute for human life. If you think colours are unreal, go and look at a painting by Bellini or Olitski, and you will change your mind. If you think reasons and values are unreal, go and make a choice, and you will change your mind.” For Korsgaard, the reflectivity of the first-personal stance cannot but raise us out of the non-personal forces that dominate the third-personal realm of explanation, at 

---

17 Korsgaard sometimes uses this phrase in a mocking sort of way, as if she were addressing those worried about a construct that is best set aside, a sort of philosophers’ bugbear. But for all that, the worries she sometimes mocks clearly preoccupy her as well.

18 Korsgaard, The Sources of Normativity, 160. Here Korsgaard goes beyond mere consistency of normativity with her materialism to a grounding relation. In my reading, she does this for good reason, seeing that for a materialist there must be some story of how normativity arose from the natural facts (since, originally at least, there were only such facts). The materialist, it would seem (or at least Korsgaard appears to hold—rightly so, in my estimation) owes some story of how this happens within the ontology she accepts, not merely a story that appeals to non-natural properties of which no materialist account is given. Be that as it may, it is clear that Korsgaard—for whatever reason—has a grounding relation in view.

At the same time, it should be noted that Korsgaard also claims that “her account does not depend on the existence of supernatural beings or non-natural facts, and it is consistent with although not part of the Scientific World View. In that sense, it is a form of naturalism,” (Korsgaard, The Sources of Normativity, 160; emphasis added, capital letters in original).

19 Korsgaard, The Sources of Normativity, 124-5.
least from the perspective of our own decision. Following what she takes to be Kant’s position, Korsgaard formulates two “standpoints” from which an action may be viewed. On the one hand, there is the third-person standpoint where explanations and predictions dominate. On the other hand, there is the standpoint of human beings as such, needing to make decisions that are justified in their own sight. “The two standpoints cannot be mixed because these two enterprises—explanation and decision—are mutually exclusive.”

The divide between these is not, for Korsgaard, to be understood ontologically; rather, “the incongruity simply follows from the fact that we stand in two very different relations to our actions: we must try to understand them, but we must also decide which ones to do.”

How we view an action depends on what we are trying to do: explain and predict, or decide and justify.

From the first-person standpoint relevant to the uniquely human business of deciding and justifying, we cannot help but view ourselves as free with respect to what we will do, and so we find that we need reasons for our choices. Our uniquely human problem is not so much that we are dominated by the forces that explain things within the Modern Scientific World View, but rather that—having somehow “stepped back” from these forces—we find no reasons addressed to us from that quarter and must now make our way with no guidance from without. This state of having stepped back from the oppressively close skein of things that dominate the other animals “sets us a problem no other animal

---

20 Korsgaard, Creating the Kingdom of Ends, 204. In both this passage and the next, Korsgaard clearly displays her Kantian heritage, repeating Kant’s own division of the cognitive business of explanation and the practical business of decision.

21 Korsgaard, Creating the Kingdom of Ends, 205. There is a strong affinity with some things Thomas Nagel says in Chapter 6 of The Last Word.
has... I desire [like other animals] and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a certain distance. Now the impulse doesn’t dominate me and now I have a problem. Shall I act? Is this desire really a reason to act? The reflective mind cannot settle for perception and desire. It needs a reason.

Having “stepped back” we find that nothing in the nexus of things from which we

22 Korsgaard, *The Sources of Normativity*, 93; emphasis in original. Paul Guyer also notes the importance of this Kantian theme in Korsgaard’s thought, agreeing with her that this is “a power that we humans have but other organisms know to us do not,” (Paul Guyer, “The Value of Reason and the Value of Freedom,” *Ethics* 109, no. 1 (October 1998): 25). Korsgaard clearly does not mean to deny that non-human animals have desires. Rather, she has in mind the uniquely human ability to “back up” from impulses and thereby gain “distance” from them, and thus have a problem no animal we are aware of has: “Should I act on this desire?”, where the “should” has the strong sense of what Korsgaard calls a “justifying reason.” The point Korsgaard is after is related to a point made very nicely by Charles Taylor as he summarizes a question Herder pursued: “What is it which makes it possible for us to have this distinct, focussed [sic] awareness of things, where animals remain caught in the dream-like, melodic flow of experience,” (Charles M. Taylor, *Hegel* (Cambridge: Cambridge University Press, 1975), 19). How, in other words, do we—alone among all creatures we know of, apparently—step back from the oppressively close wash of experiences and desires such that we are aware of having options about what to do and need reasons for taking one course of action over another?

The language of “stepping back” or gaining “distance” also appears prominently in the work of Thomas Nagel and John McDowell when they turn their attention to ethics. See Nagel, *The View from Nowhere*, 138-40, 186-87, and John McDowell, “Two Sorts of Naturalism,” in *Mind, Value and Reality* (Cambridge, MA: Harvard University Press), 171-74, 188-89. In each case, it strikes me as deeply problematic, as they essentially grant themselves an answer to some of the most difficult problems surrounding the mind and meaning. Under cover of a metaphor of “stepping back,” they help themselves to an answer to one of the most difficult problems in human thought, one that Martin Heidegger spent a lifetime thinking about and never really resolved: how did we, alone among the animals, emerge from the oppressively close nexus of things? As Heidegger made the point, “Da-sein means: being held out into the nothing. Holding itself out into the nothing, Dasein is in each case already beyond beings as a whole [i.e., the nexus of things]. Such being beyond beings we call *transcendence*. If in the ground of its essence Dasein were not transcending, which now means, if it were not in advance holding itself out into the nothing, then it could never adopt a stance toward beings nor even toward itself. Without the original manifestness of the nothing, no selfhood and no freedom,” (Martin Heidegger, “What Is Metaphysics?” in *Pathmarks*, ed. William McNeill (New York: Cambridge University Press, 1998), 91.). Or, as Korsgaard herself quotes Kant as saying: The first human “discovered in himself a power of choosing for himself a way of life, of not being bound without alternative to a single way, like the animals… He stood, as it were, at the brink of an abyss,” (Kant, “Conjectural Beginnings of Human History,” trans. Emil L. Fackenheim, in *Kant on History*, ed. Lewis White Beck, Library of Liberal Arts (1786; New York: Macmillan, 1963), 111-12, quoted in Korsgaard, *Creating the Kingdom of Ends*, 112).

Heidegger thought much more deeply about the implications of the abyss opened up by this “step back” from or (as he would have put it, as “stepping back” already subtly grants the subjectivity in question) “falling away” of things than Korsgaard, Nagel, McDowell, and other contemporary Kantian analytic philosophers. Not to put too fine a point on things, it is often remarked that Heidegger did not develop any moral philosophy. Though this is not the place to do so, I would argue that (at least part of) the reason for this is because he thought much more deeply about what Korsgaard calls “the plight of rational agency” (Korsgaard, *Creating the Kingdom of Ends*, ii)—this “step back” from or “falling away” of things—than Korsgaard, Nagel, and McDowell. Thus, for a giraffe, her “perceptions are the ground of her actions. Because she sees the world that way, with the reasons already loaded into it, she nearly always already knows what to do”
have somehow gained distance addresses a reason to us. An apparent implication of the Modern Scientific World View is that when we step back from the nexus of things, there is nothing “out there” of the right kind to address reasons to us, nothing that has moral value apart from our act of conferring value through acts of endorsement. And so we human beings must exercise our “creative power with respect to the making of values.”

We step back and find ourselves alone, with no moral reasons or moral values but those which we confer by endorsing a maxim. So far as we know, we are the only beings in the universe with the capacity to imbue things with moral value and to give to ourselves the reasons that justify our actions. Alone of all that we know of, “we confer value on the objects of our rational choices.”

Moral value, at least, is conferred by humans as we endorse an end as

(Korsgaard, *Self-Constitution*, 212). Things are different for us: “Self-consciousness… separates your perceptions from their automatic normative force” (213). With normativity having “fallen away,” we are in a plight where we must decide what to do without guidance from a world that is normativity. This “plight” is deeper than Korsgaard realizes, though it has assumed an increasingly prominent place in her thought of late (witness the title of her two most recent books: *The Constitution of Agency* and *Self-Constitution*). It would be a fascinating and potentially important study to explore Heidegger’s insights and the reasons for his failure to develop any moral philosophy despite his sustained, intense interest in many of the same philosophers Korsgaard draws on in formulating her own moral philosophy—including Kant and Aristotle. This dissertation, however, must largely pass by such explorations.


24 Korsgaard, *Creating the Kingdom of Ends*, ix; emphasis in original. In an interesting review of *Creating the Kingdom of Ends*, it begins to dawn on Allen Wood that Korsgaard’s understanding of “conferring value” differs from the sense in which a university might confer a degree, the latter sense having to do with a recognition of requirements that are met. “At times it seems as though she understands the conferral of value to amount to no more than a declaration that a thing (or a proposed end) should be called ’good’ and made the object of a certain behavior on the part of rational agents,” (Allen W. Wood, Review of *Creating the Kingdom of Ends*, *The Philosophical Review* 107, no. 4 (October 1998): 610). As Allan Gibbard clearly sees, endorsement really is central to Korsgaard’s moral philosophy. This also emerges very clearly in a thoughtful reading of *The Sources of Normativity* in Michael E. Bratman, “The Sources of Normativity,” *Philosophy and Phenomenological Research* 58, no. 3 (September 1998): 699-709.
worthy of being acted upon.25 Korsgaard’s Kant is much like the metaphysically thin Kant of John Rawls’ Dewey Lectures.26

This “power to set an end (to make something an end by conferring the status of goodness on it)” Korsgaard follows Kant in calling “humanity.”27 “When Kant says that the characteristic of humanity is the power to set an end… he is referring to a more general capacity for choosing, desiring, or valuing ends; ends different from the ones that instinct lays down for us, and to which our interest is directed by the operations of reason.”28 Korsgaard identifies Kant’s “crucial step” in his formulation of the categorical imperative as the supposition “that rational choice itself makes its object good.”29 This is part and parcel

---

25 I believe Korsgaard holds this to be the case with value generally, not just moral value, but that is of no concern here. Korsgaard would take herself to be following the famous opening line of Kant’s *Groundwork*, where he emphasizes that only the “good will” can be considered “good without limitation.”

If human beings have an intrinsic value by virtue of the capacity for valuing things, then human beings bring goodness into the world. The distinction between a thing that is intrinsically good and a thing that is extrinsically good yet valuable as an end allows for the possibility that the things that are important to us have an objective value, yet have that value because they are important to us. Objective goodness is not a mysterious ontological attribute. The things that are important to us can be good: good because of our desires and interests and loves and because of the physiological, psychological, economic, historical, symbolic and other conditions under which human beings live, (Korsgaard, *Creating the Kingdom of Ends*, 273; emphasis added).


27 Korsgaard, *Creating the Kingdom of Ends*, 124. See also, Korsgaard, *Creating the Kingdom of Ends*, 137, 260. A passage from Kant that Korsgaard quotes in support of her position is quoted at note 103 below. A number of commentators have noted the central importance of this idea of “humanity” for Korsgaard’s moral philosophy. See Guyer, “The Value of Reason and the Value of Freedom,” 25-6; Hannah Ginsborg, “Korsgaard on Choosing Nonmoral Ends,” *Ethics* 109, no. 1 (October 1998): 6; Wood, “Review of *Creating the Kingdom of Ends*,” 609.

When in the text below “humanity” or the “human” is referred to in connection with Korsgaard’s moral philosophy, it should be understood in Korsgaard’s special sense of these words.

28 Korsgaard, *Creating the Kingdom of Ends*, 114.

29 Korsgaard, *Creating the Kingdom of Ends*, 122; emphasis in original. See also Korsgaard, *The Sources of Normativity*, 120-23.
of what Korsgaard calls the “reflective endorsement method” of establishing normativity.\textsuperscript{30} In \textit{The Sources of Normativity}, Korsgaard first introduces this notion in connection with David Hume, and notes that this method “has its natural home in theories that reject realism and ground morality in human nature.”\textsuperscript{31} For Korsgaard, as for David Hume, human morality arises from “a productive faculty” in which our activity of “gilding or staining all natural objects with the colours, borrowed from internal sentiment, raises in a manner a new creation.”\textsuperscript{32} Our endorsement of an end \textit{just is} what makes it worth pursuing, and apart from such endorsement, there would be neither values, nor ends, nor reasons. No endorsement, no values. This is made clear in Korsgaard’s response to the case of an “idealized Mafioso” that G.A. Cohen poses as a problem.\textsuperscript{33} Doesn’t the Mafioso have reasons to murder people obstructing the Family emerging from what he endorses from within his “practical identity”? Korsgaard’s response is clear: “There is a sense in which these obligations [stemming from the Mafioso practical identity] are real—not just psychologically but normatively. And this is because it is the endorsement, not the explanations and arguments that provide the material for the endorsement, that does the normative work.”\textsuperscript{34} Endorsement is the engine driving valuation.

\textsuperscript{30} Korsgaard, \textit{The Sources of Normativity}, 50.

\textsuperscript{31} Korsgaard, \textit{The Sources of Normativity}, 50.

\textsuperscript{32} Hume, \textit{An Enquiry concerning the Principles of Morals}, Appendix I.

\textsuperscript{33} The “idealized Mafioso” will be considered more fully below.

\textsuperscript{34} Korsgaard, \textit{The Sources of Normativity}, 257. “It is necessary to have \textit{some} conception of your practical identity, for without it you cannot have reasons to act. We endorse or reject our impulses by determining whether they are consistent with the ways in which we identify ourselves,” (Korsgaard, \textit{The Sources of Normativity}, 121; emphasis in original). As even the title of her book suggests, it is endorsement from within practical identities that is the engine of normativity. These identities are the \textit{sources} of normativity.
But while endorsement is central for morality, Korsgaard argues that the very nature of human endorsement pushes us beyond Hume and his epigones, to Kant. Specifically human endorsement—endorsement by a human qua human—is reflective endorsement, and this is Kant’s fundamental insight, in Korsgaard’s view. Not just any endorsement rightly creates value and sets an end, but only endorsement that expresses rational reflection.  

“If you had no normative conception of your identity, you could have no reasons for action, and because your consciousness is reflective, you could then not act at all. Since you cannot act without reasons and your humanity is the source of your reasons, you must value your own humanity if you are to act at all.”

Moral action, action that is genuinely human, is a product of endorsement that expresses our autonomy as rational beings who give ourselves our own law. This is our power of rational choice, the very power that allows us to “step back” from the oppressively close skein of things in which the non-rational animals are submerged. The very act of deciding what to do is itself expressive of this capacity and rational agents cannot help but performatively accept its value; as humans who must decide what to do, we value this capacity in the very act of its exercise. “If complete normative skepticism is to be avoided—if there is such a thing as a reason for action—then humanity,

---

35 As noted above (pp. 291-94), “humanity” here should be understood in Korsgaard’s special sense. She certainly does not mean to say that all homo sapiens always act reflectively—though she would say that they are not acting from out of their humanity when they are swept along merely by desire. Korsgaard’s Kantianism is clearly on display here, and one can see an especially close connection with what Kant says in the *Groundwork* about “pathological” acts in which we let desire dominate our choice and thus forego the one thing that is essentially human—autonomous choice. Failing to give a law to ourselves, the world legislates over us, thereby reducing us to something less than truly “human.”

36 Korsgaard, *The Sources of Normativity*, 123.

37 Korsgaard, *Creating the Kingdom of Ends*, 167.
as the source of all reasons and values, must be valued for its own sake.”\textsuperscript{38} The only demand the capacity for reflective endorsement places upon us is that we act consistently with respect to this capacity itself: performatively, we cannot overturn its value, nor can we will in contradiction to our own decision. Doing so would be to denigrate the value that we have ourselves set, a conflict of valuations that cannot be sustained. Resolution must be sought, regressing upon the conditions of our own willings until we come to their source and the one thing that carries its value for us in itself: “It is our capacity to set ends—to freely choose what shall be an end by means of reason, that not only makes every rational being an end in itself, but which forms the only possible final purpose of nature, teleologically conceived. It is only this capacity that has its value completely in itself.”\textsuperscript{39} Any action—as opposed to being carried along by inclination and incentive as if we were mere animals (a “pathological” happening, as Kant would call it)—presupposes this value of humanity, that is, human autonomy as beings who decide what to do from within a standpoint of spontaneity.\textsuperscript{40}

Autonomous action flowing from a rational will is not random, but is action according to law; indeed, it emerges from lawfulness itself. As Korsgaard puts it, spontaneous human willing has “no constraint on its choice, except that it choose a law.”

\textsuperscript{38} Korsgaard, \textit{The Sources of Normativity}, 122.

\textsuperscript{39} Korsgaard, \textit{Creating the Kingdom of Ends}, 130.

\textsuperscript{40} Kant is particularly concerned to show that free actions are not arbitrary, but are instances of lawfulness. He is responding to the sort of criticism offered by John Locke: “A perfect indeterminacy in the mind, not determinable by its last judgment of the good or evil that is thought to attend its choice, would be so far from being an advantage and excellency of any intellectual nature, that it would as great an imperfection, as the want of indifferency to act, or not to act, till determined by the will would be an imperfection on the other side,” (John Locke, \textit{An Essay Concerning Human Understanding}, vol. 1, ed. Alexander Campbell Fraser (1690; New York: Dover, 1959) II, xxi, 49 (also 50-51)).
The only constraint on spontaneous willing is the internal requirement constitutive of spontaneity itself, namely that the will “have the form of a law. Nothing provides any content for that law; all that it has to be is a law. The moral law simply describes the position of a free will. When the will’s choices are directed by the moral law, it expresses its spontaneity. The moral law is the law of spontaneity.” It is nothing but the will to act consistently with a law rather than being swept along by inclinations, and as such, our decisions from within the standpoint of our own spontaneity (that is, our capacity to “step back”) must do nothing else but will the consistency of our will with itself. Only this is freedom from inclination. So the “only constraint on our choice is that it have the form of a law. Nothing provides any content for that law. All that it has to be is a law.” This, again, just is what it is for the free will to be a free will; it “simply describes the function or task of an autonomous will.” In a real sense, we can never choose against the categorical imperative—we can only be submerged again in brute animality and carried off, “pathologically” dominated by heteronomous desires.

Korsgaard, however, is not fully satisfied with this account. She suggests that the categorical imperative is not sufficient of itself to vouchsafe the moral law, for it does not settle the question of the domain over which the free will ranges. “It is only if the law ranges over every rational being that the resulting law will be the moral law.”

---

41 Korsgaard, *Creating the Kingdom of Ends*, 25. Making much the same point in *The Sources of Normativity*, Korsgaard states that “autonomous lawmaking just isn’t autonomous lawmaking unless it is done universally. The requirement of universalization is not imposed on the activity of autonomous lawmaking by reason from outside, but is constitutive of the activity itself,” (235; emphasis in original. See also, p. 98).

42 Korsgaard, *Creating the Kingdom of Ends*, 166; emphasis in original.

43 Korsgaard, *Creating the Kingdom of Ends*, 166.

admits that Kant himself thought that the categorical imperative was sufficient to give us the moral law of itself, and some commentators think Korsgaard might have been better served by hewing more closely to what they take to be Kant’s own view. In any case, the objection Korsgaard has in mind is the classic objection that the egoist may admit that she must value her own humanity, while finding no reason to continue on from there to find value in the humanity of others. As Raymond Geuss puts it, “I may well come to see my mere humanity as a source of value for me, your mere humanity as a source of value for you; how does it follow from that that your humanity must be a source of value for me?” Again, what is the domain over which my valuing of humanity must range? Must I value it only in myself, or must I value it in others as well?

B. Identities and Maxims: Projecting the Logos

At the center of Korsgaard’s attempt to overcome the flaw she sees in grounding moral philosophy in the categorical imperative alone is her account of “practical identities.” In *The Sources of Normativity*, Korsgaard identifies the practical identities of human beings as the sources of normativity, including values and obligations. A practical identity is essentially a role that you accept (whether rationally or not, as her response to the Mafioso

---

45 In the final analysis, Korsgaard does appear to circle back to the categorical imperative to do the work of getting from identities to moral law—or, at any rate, the answer turns out to be that reflection limits the identities that one can accept as fully consistent with one’s own humanity. As Schneewind notes, if in the final analysis the reason why not just any identity can be accepted is that it is “simply inconsistent” to do so, “then she seems to belie her own claim that the CI [categorical imperative] alone does not get the ML [moral law].” (J.B. Schneewind, Review of *The Sources of Normativity*, *Mind* 106, no. 424 (October 1997): 794). Bratman, “The Sources of Normativity,” 703, makes a similar observation.

problem makes clear), an understanding of yourself that you regard as having value. Your practical identity is “a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking.” Practical identities range across family ties (being Russ’ son or Emma’s mother), love and friendship (being Kiersten’s husband or Laura’s friend), professional identities (being a professor or a hairdresser), religion (being Christian or Atheist), defining interests (being a volleyball player or a cellist), and so on. But taking up any of these identities performatively requires valuing your own “humanity.” None of these identities is to be understood along “theoretical” lines as explaining the scientific facts about you, but rather from the standpoint of making decisions, of spontaneity. From a third-person, explanatory standpoint, you might be explained as a phenotypic vehicle for your genotype, and Korsgaard, as a materialist, is committed to some such explanation of human beings. But Korsgaard wants to focus on our practical identity, not our theoretical identity. Your practical identity is an understanding of yourself that you endorse, and, as such, out of which you live daily life, making decisions expressive of value ranging from whom you will befriend or marry to how you will spend your time and money.

---

47 Korsgaard, *The Sources of Normativity*, 101. As should already be clear from what we have seen above (and will be developed more fully below), there is more to morality than just “practical identities,” and rational constraint of these identities is critically important in that overall picture, but practical identities themselves—including their normative force—need not be rational. See again Korsgaard’s response to Cohen’s Mafioso challenge (p. 292) above.

48 Korsgaard is not meaning to deny that there may often be distinctions between one’s actual practical identity and one’s accepted or nominal practical identity. She need not rule out Freudian, Marxist, or Pauline hermeneutics of suspicion. However, the point she is making here is not directly affected by this distinction, though the specific outworking of her views in actual human moral decision making would need to be sensitive to such issues.
Identities explain why we have the reasons we do. As Russ’ son I have reasons to drive back to Gurnee, Illinois that I would not have if I were the son of Nate, next door in Upland, Indiana. As a cellist you have reasons for spending your time in particular ways that I do not have, say, practicing a piece of music for two hours. Our endorsement of our practical identities makes our decisions comprehensible from the standpoint of an agent who acts. Practical identities “give rise to reasons and obligations. Your reasons express your identity, your nature; your obligations spring from what that identity forbids.”

Who you take yourself to be leads you to endorse certain maxims: “[Since I am Russ’ faithful son, and] since Russ is sick in the hospital, I endorse the maxim, ‘I will go to the hospital to love my Dad’.” If, as Korsgaard puts it, “normativity is built right into the role,” then it seems we have the beginnings of a way forward with one of the most difficult problems for moral philosophy within the Modern Scientific World View. “I must take some ways of identifying myself seriously, or I won’t have any reasons at all. Insofar as I take that fact—the fact that I need some way of identifying myself—to be a reason, I express the value I set on myself as a human and rational being.”

There is no need to ground normativity in any queer entities or properties, or for persons accoutered with the capacity to perceive them somehow. Normativity is grounded in our practical identities, and reasons are

49 Korsgaard, *The Sources of Normativity*, 101. One of the challenges Korsgaard will need to meet is, as Paul Moser has commented (personal communication, 6 June 2008), why we shouldn’t “insert ‘supposed’ in front of: reasons, identity, nature, and obligations.” Korsgaard will attempt to meet this challenge by appealing to the rational nature we cannot help but appeal to in taking up any practical identity. This response is developed below, and its adequacy questioned.

50 Thus the sense of practical identities “giving rise to reasons and obligations” seems to be related to a kind of motivational internalism. This accords with Gibbard’s (what seems to me correct) claim that Korsgaard has some fundamental affinities with expressivism, laying a huge burden on endorsement in the genesis of normativity. Again, to see this role laid out very clearly, see Korsgaard’s response to Cohen’s Mafioso problem (as summarized on pp.271-2 above).

expressive of the valuations we make within those identities. Neither are obligations mysterious in the slightest. Rather, they can be seen to be actions that void our identities if we take them or fail to take them—things we can do or fail to do only at the cost of losing ourselves. “An obligation always takes the form of a reaction against a threat of a loss of identity.”52 Since we understand the notion of having an identity well enough, the problems with understanding obligation and normativity generally can seem tractable if they can be understood to be tied in this way to our practical identities.

The medium through which our practical identities operate is what Korsgaard calls a maxim. As we reflect about what to do within the practical standpoint, where we must decide what to do based on our practical identity, what we consider are maxims. Korsgaard says a number of different things about maxims in various places in her writings, and it will be argued below that they are of dubious self-consistency. Nevertheless, a picture of what we are doing in reflection begins to emerge. Korsgaard sketches a picture of reflective endorsement of a maxim in a passage worth quoting at length:

Roughly speaking, what happens when an agent chooses an action is something like this: The agent is attracted on some occasion to promoting some end or other. The end may be suggested by the occasion, or it may be one he standardly promotes when he can. He reasons about how he might achieve this end, or what he might do in its service, and he arrives at a possible maxim or logos. He considers promoting a certain end by means of a certain act done in a certain way at a certain time and place. That is to say, he considers an action, and he asks himself whether it is a thing worth doing.53

---

52 Korsgaard, _The Sources of Normativity_, 102.

Our reasoning produces a maxim or *logos*, where this is understood to be “both the act done and the end for the sake of which that act is done.” According to Korsgaard, the “basic form” of a maxim is, “I will do act-A in order to promote end-E.” The maxim is at the heart of Korsgaard’s moral philosophy, for it is here that human beings propose ends for which an action may be done. Thus, when the agent acts on a maxim, he “chooses not only the act, but the purpose or end—he chooses the act for the sake of the end, but in doing so he chooses to promote or realize the end.” Our act of endorsing a maxim sets an end in place, the very declaration of value itself conferring the value. “Ends are chosen as parts of maxims which in turn are chosen as laws the agent gives to herself.” The process of reasoning that goes into the formulation of a maxim proposes an end that will be the guide for the proposed act. When an agent acts in accord with a maxim, value is conferred and an end is set in place.

It is at this point that there is a tension in Korsgaard’s thought about maxims. The tension has to do with the relationship between maxims and reflective endorsement. On the one hand, in the passage quoted at length above, the maxim appears to be a *candidate* for

---


55 Korsgaard, “Acting for a Reason,” 22. See also Korsgaard, *Creating the Kingdom of Ends*, 57-8.


58 Even the value of being a being who chooses is conferred by our choosing: performatively, when I make any choice, I cannot help but value my ability to choose, thus conferring value upon it. Korsgaard’s view of the role of humanity in conferring value goes quite deep: “I am prepared to agree that if human beings decided that human life was worthless then it would be worthless… it is the endorsement that does the work,” (Korsgaard, *The Sources of Normativity*, 254; emphasis in original).
reflective endorsement (although Korsgaard does state that it is a possible maxim). And in *The Sources of Normativity*, Korsgaard speaks of good and bad maxims, implying that one must somehow distinguish between them by appeal to another step. It is the reflective endorsement test that fills this role.\(^{59}\) Reflective endorsement here appears as an additional step beyond the formulation of a maxim that elevates some maxims to a status in which they become reasons for action. Here your maxim “expresses what you take to be a reason for action.”\(^{60}\) On this understanding, the maxim is merely a candidate for ratification by the reflective endorsement test—a test that finally appeals to an overarching identity shared by all human beings as such. Such maxims-as-candidates for law may thus be called identity-priority maxims, and the role of the overarching human identity in this picture will be explored below.

On the other hand, one sometimes finds Korsgaard understanding maxims very differently, apparently taking them to be sufficient in themselves to justify an action. On this view, a maxim does not need anything else in order to be a reason for action; no additional step of reflective endorsement is needed. Rather the maxim *at such* embodies or expresses reflective endorsement due to the very nature of the act by which it is formulated. Thus Korsgaard can hold that her account of maxims “correctly identifies the kind of item that can serve as a reason for action: the maxim *or* logos of an action, which expresses the agent’s endorsement of the appropriateness of doing a certain act for the sake of a certain end.”\(^{61}\)

\(^{59}\) Korsgaard, *The Sources of Normativity*, 108.

\(^{60}\) Korsgaard, *Creating the Kingdom of Ends*, 57.

\(^{61}\) Korsgaard, “Acting for a Reason,” 28. It should be noted that at times in this important essay, Korsgaard speaks of the “categorical imperative test” that a maxim must pass (21). It’s unclear how this relates to her understanding in *Sources* that the categorical imperative falls short of vouchsafing the moral law.
this understanding, the maxim itself expresses endorsement of an act-end pair by an agent, thus expressing the agent’s reason for acting. The key here is that it is the form of a maxim that makes it what it is. “The maxim of the action must be legal in itself, and this can only be because it has a lawlike form.”62 In laying out her understanding of a “good maxim,” Korsgaard speaks of it as being good “in virtue of its internal structure… A good maxim is therefore an intrinsically normative entity. . . . an ‘entity’ whose intrinsic properties, or internal structure, renders it normative.”63 She goes on to say that this renders realism true after all, for there is something about the nature of the good maxim in itself that makes it intrinsically normative. What does the work here is the form of the maxim itself; a maxim that has lawlike form expresses what truly is a reason to act—and this because of the form that is intrinsic to it. Korsgaard’s most recent book advances the point with startling clarity: “the maxim itself, and therefore the action, has the property of lawfulness. What is the property of lawfulness? A law lays down what is to be done, so we might say it is to-be-doneness, that famous property that John Mackie claimed nothing in the world as we know it could possibly have.”64 Understood this second way, maxims may be called form-priority maxims.

This second reading of Korsgaard on maxims is made all the more interesting because of her identification of the maxim and the logos. In two of her most recent reflections on moral philosophy she lingers over the idea of the logos, four times explicitly identifying the logos and the maxim in her essay “Acting for a Reason.”65 In another piece,

---

62 Korsgaard, Creating the Kingdom of Ends, 62.
63 Korsgaard, The Sources of Normativity, 108-9; emphasis in original.
64 Korsgaard, Self-Constitution, 16.
written shortly before “Acting for a Reason,” Korsgaard reflects extensively on the notion of the *logos* and its connection with human nature. Here she speaks of the *logos* as being “deeply embedded in human nature, so deeply that it is difficult even to talk about, and it stands behind the whole range of our values.” The *logos*, in this essay, is also understood as a ground or as the idea of a story, and it seems to be the basic structure of justification and understanding itself. “One thing does not merely follow another: it justifies it, explains it, rewards it, punishes it, it is its climax, its culmination, its fruition, or its doom.” Now, what Korsgaard has in mind here is not wholly clear. But she seems to be saying that the *logos* is the narrative structure itself; the following of one idea upon another—the basic normativity embedded even in language itself as the meaning that leads us from one idea to another and weaves them together in support relations. “It is a primitive form of value, inhabiting all others, a kind of general structure of value.” Now, if Korsgaard would carry over what she says here about the *logos* to her later essay where she identifies the *logos* and the maxim, this would leave us with a very strong reading of the centrality of the maxim for

---

66 Korsgaard is somewhat apologetic when speaking of human nature, saying that it is “an old-fashioned item,” (Christine M. Korsgaard, “The Dependence of Value on Humanity,” in Joseph Raz, *The Practice of Value* (Oxford: Clarendon Press, 2003), 65). After expressing her slight embarrassment, however, she just dives right in and talks about it quite a bit indeed.

I find it particularly interesting that Korsgaard’s reflections about the *logos* occur within the context of thinking through Joseph Raz’s illustration of one of his points by reference to architecture. Korsgaard draws out this example over a few pages, drawing connections between the practice of architecture—and particularly the boundaries inherent in such practice—and the practice of ethics. I have explored these connections previously in my Masters Paper written for Andrew Cutrofello (Bradley N. Seeman, “Unsettling Deconstruction: Architecture as Humanistic Frame,” unpublished mss. (October, 2006)). The connections between Korsgaard and Derrida on the role of place, boundary, and *logos* in moral philosophy would merit further exploration.


69 Korsgaard, “The Dependence of Value on Humanity,” 72. This seems to me to be an incipient Heideggerian movement. Here, again, it would be interesting to compare what Korsgaard says here with what Derrida says about *differance/chora.*
moral philosophy. On this very robust reading, the maxim/logos structure would be the primitive normativity that underlies all value, all meaning. It would make humanity, in Korsgaard’s sense, possible—the logos/maxim-structure calling out the human identity. At the heart of making any sense of our experience is the maxim/logos. Somehow, there is a fundamental value that is embedded in the maxim itself—that just is the maxim, in a sense. And it is only in light of what is there in the maxim that any affirmation as such can arise in the first place. In that case, the maxim itself is the uniquely human “stepping back” from being submerged in the skein of causality along with all non-rational animals. This fits very well with the arc of Korsgaard’s thought about human identity as constructed (as will be seen below). In light of this, in those places where some of what Korsgaard writes hints at the form-priority maxim view, she may be thinking of the maxim as the expression of freedom itself, the shape of the clearing in which we think, the very possibility of any identity whatsoever—with all identities owing to the form of the maxim.

If it is correct that there are two plausible readings of Korsgaard on maxims, either as identity-priority maxims or as form-priority maxims, then there is a difficulty of whether these can be harmonized, and, if not, which reading should be preferred and for what reasons. The first reading would be closer to Kant’s own understanding of maxims, as many

70 Once again, this would make for potentially fruitful comparisons between what Korsgaard has to say here and what Derrida (chora) and Heidegger (das Nichts) have to say. And, again, given such a connection and the common metaphysical commitments of these thinkers, I would tend to say that Heidegger’s silence regarding moral philosophy or Derrida’s (on my strong reading of the implications of chora) struggles to allow “the other” his or her own voice would be more consistent than Korsgaard’s attempt to formulate a fairly strong Kantian moral philosophy on these grounds. See note 22 above.

71 I want to make clear that I am not claiming that the reading I have just given is something Korsgaard herself necessarily has in mind (though, as mentioned above, her latest work has moved in this direction). Rather, the reading I have given is a proleptic attempt to bring out a germ that I see in her latest writings that might represent a direction a Korsgaard-inspired moral philosopher might try develop from out of Korsgaard’s writings. I will attempt below to show that such a move (whether made by Korsgaard or another) would not serve the aim of giving authoritatively prescriptive moral reasons that are needed if one is to avoid moral reasons arbitrariness (as Korsgaard clearly hopes to do).
commentators understand him (although one could certainly find the spirit of the second reading in some things Kant says) and this seems to be Korsgaard’s official stance. For reasons that will appear shortly, however, it seems that Korsgaard’s position is shifting toward the latter position. Indeed, in Korsgaard’s latest book the latter position seems to dominate.

It is of central importance for understanding Korsgaard on these issues to see why, in *The Sources of Normativity*, she sees that what are here being called form-priority maxims are insufficient for giving the moral law—and why she might be backtracking in her latest work. In summing up the core of her argument in her pivotal third lecture in *Sources*, Korsgaard says that maxims get us only so far: “When an impulse—say a desire—presents itself to us, we ask whether it could be a reason. We answer that question by seeing whether the maxim of acting on it can be willed as a law by a being with the identity in question. If it can be willed as a law it is a reason, for it has an intrinsically normative structure. If it cannot be willed as a law, we must reject it, and in that case we get an obligation.”

This is the end-result of the test of reflective endorsement applied to maxims (where the official reading of these—although elements of the other reading are clearly beneath the surface—is definitely the identity-priority maxims reading). Some maxims—the “good” ones—pass this test and are endorsed by the agent’s will, thus giving us a *reason*. Maxims that do not pass the test are

---

72 Korsgaard, *The Sources of Normativity*, 113. It is interesting that in this summary we find one of Korsgaard’s Heideggerian *Seinsfrage*-flavored passages: “The reflective structure of human consciousness sets us a problem. Reflective distance from our impulses makes it both possible and necessary to decide which ones we will act on: it forces us to act for reasons” (113).

73 Note that the structure of the preceding passage runs the core of these two readings together. Is the willing what gives reasons and obligations, or is the structure of the maxim what gives reasons and obligations? On the reading advanced here, this verbal tension is not accidental, but is indicative of a structural fault-line running through the center of Korsgaard’s work.
rejected, giving us an obligation to avoid the action-end pair in question. The maxim itself must be ratified by the reflective endorsement of the agent from within the identity from which she acts.

But Korsgaard remains worried. She sees that “there is still a deep element of relativism in the system. For whether a maxim can serve as a law still depends upon the way that we think of our identities.” Given some identities, someone might will all manner of atrocities quite consistently with that identity. Practical identities, after all, come not merely in the form of wives, children, stamp-collectors, and shop-keepers, but also in the form of dictators, corrupt businessmen, Mafiosi, and the like. Faced with such difficulties—sometimes pushed very forcefully, as with G.A. Cohen’s Mafioso example (explored below)—Korsgaard works to “establish that there are particular ways in which we must think of our identities.” It is reflection that leads us out of this difficulty, according to Korsgaard; for the activity of reflection “has rules of its own… And one of them, perhaps the most essential, is the rule that we should never stop reflecting until we have reached a satisfactory answer, one that admits of no further question. It is the rule, in Kant’s language, that we should seek the unconditioned.” As Korsgaard puts it elsewhere, “if the value of something is conditional… an inquiry into the conditions of its value should lead us eventually to what is unconditioned.” Such reflection leads finally to the good will as the

74 Korsgaard, The Sources of Normativity, 113.
75 Korsgaard, The Sources of Normativity, 113.
76 Korsgaard, The Sources of Normativity, 258.
77 Korsgaard, Creating the Kingdom of Ends, 118.
“unconditioned condition of the goodness of other things.” Through such reflection, Korsgaard argues, the person who is capable of adopting any practical identity at all is led back to a generic human identity that we cannot help but have, and that this human identity is at the root of reflective endorsement, thus limiting the ways in which reflective endorsement can operate from out of any identity whatsoever. If “your reasons express your identity,” then what is needed is a way of limiting the identity out of which reflective endorsement occurs.

To make a long story short, Korsgaard argues that having any practical identity at all—and one cannot help but have some such identity insofar as one acts at all—is to value the reflective endorsement that opens up the possibility of any practical identity whatsoever.

Most of the time, our reasons for action spring from our more contingent and local identities. But part of the normative force of those reasons springs from the value we place on ourselves as human beings who need such identities. In this way, all value depends on the value of humanity; other forms of practical identity matter in part because humanity requires them. Moral identity and the obligations it carries with it are therefore inescapable and pervasive. Not every form of practical identity is contingent or relative after all: moral identity is necessary.

In acting, you act from the kind of endorsement characteristic of humanity (in Korsgaard’s sense) as such—namely, the identity of a being who chooses after reflecting on the matter—and thus performatively demonstrate your valuing of this capacity that makes humans human. But acting from this capacity identifies you as a human, who cannot value any practical identity at all without valuing the humanity that opens up the possibility of that identity, a humanity that you share with all human beings generally. Your valuing of humanity in

---

78 Korsgaard, *Creating the Kingdom of Ends*, 118.


yourself is a valuing of humanity generally, and that is shared by all other human beings—and, apparently, only human beings. In valuing any identity enough to act from out of it, then, you necessarily value the same humanity that other people act from in their own practical identities. Such is the basic idea that Korsgaard develops as the “additional step” beyond the identity-priority maxim account of moral law that she sketched earlier in the third lecture of The Sources of Normativity. Reflective endorsement by the will of human beings as such is central to this account, and some such account flowing out of the will of humans as humans would seem to be required if morality originates when maxims are ratified by reflective endorsement and thus made into genuine reasons for action (as on the identity-priority maxims account).

Human freedom in reflectivity—the “step back” from the nexus of things that dominated our primate precursors—is clearly the crucial starting point for this argument. The logic of this starting point and the criticisms to which Korsgaard opens herself as a materialist by simply granting herself the human subject at the outset puts pressure on her to adopt an alternative reading of maxims: the form-priority maxim account. Somehow one primate “steps back” from the oppressively close skein of efficient causality that heretofore

---

81 In her essay, on “Acting for a Reason,” Korsgaard speaks of a lioness protecting her cubs from “a marauding male lion” as “an agent, not a mechanism,” acting “in order to protect her cubs.” One “may even be tempted to say that she acts under the influence of a normative claim,” (Korsgaard, The Constitution of Agency, 213; Korsgaard’s emphasis). At the same time, Korsgaard emphasizes that there is “something different in the human case”: “The human being is aware of the reason as a reason; she identifies the good-making properties of the action under the description ‘good’ or ‘reason’ or ‘right,’ or some such normative description. She does not act merely in accordance with a normative consideration, but on one,” (213-14; emphasis in original). Thus, while Korsgaard seems open to something like a sort of “reflection” in higher, non-rational animals, she also emphasizes what she takes to be a key difference: the ability that humans alone have to appreciate and evaluate reasons as such—critically examining the support relations in view and acting specifically on those support relations that they endorse after reflecting on them.

82 The pivotal passages for the summary in this paragraph may be found in Korsgaard, The Sources of Normativity, 119-23.
dominated all other animals, such that none of them could appreciate, weigh, and act on the normative support relations essential to moral reasons. But, given Korsgaard’s acceptance of the Modern Scientific World View as the accurate *third-person, explanatory account*, the theoretical possibilities for this “step back” have implications for what can plausibly be said of things in the *first-person, justificatory standpoint* from which the agent makes decisions. In particular, one cannot just start off with identity, but rather there must be some sort of event that happens to the precursor primates that gives rise to the willing subject, and an event tied to the maxim/logos on the second reading seems as good as any. There was no willing, human subject prior to the “step back,” and the maxim/logos on the second reading is the structure that gives rise to the subject. As Korsgaard puts the matter in an important passage,

> The idea that reasons are private, that they might belong to one person in particular, seems to assume that we can identify one person in particular in advance or prior to the person’s reasons themselves. I also hold a kind of constructive view of personal identity, and I think that what gives us a personal identity is the reasons that we autonomously adopt for ourselves. And that means that you cannot take for granted the notion of somebody’s being me in advance of which reasons I have and adopt. My reasons are part of my practical identity, and I construct my practical identity, and so I do not exist, so to speak, ahead of my reasons.  

---

83 See note 22 above. See also, Korsgaard, Self-Constitution, 19: “The identity of a person, or an agent, is not the same as the identity of the human animal on whom the person normally supervenes… we are conscious of the grounds on which we act, and therefore are in control of them. When you are aware that you are tempted, say, to do a certain action because you are experiencing a certain desire, you can step back from that connection and reflect on it.” This begs for an explanation of how the word “normally” (what is going on when it doesn’t supervene?) and also the “you” that steps back fit into the materialist world of the Modern Scientific World View.

84 Korsgaard, “Internalism and the Sources of Normativity,” 54-55; emphasis added, with the underlining pointing out what I take to be perhaps the most important phrase for the form-priority reading of Korsgaard on maxims. Even in this important passage for the form-priority maxim view, there is obviously still a strong element of the identity-priority maxim view, for Korsgaard says it is we who autonomously adopt the reasons that give us a personal identity. I believe there is a deeply problematic circularity here having to do with whether reasons precede the self or the self precedes reasons that is symptomatic of the identity/reasons problem I will be pressing in the next section of this chapter. Moreover, I believe this circularity is a deeply rooted Kantian problem, as I shall (somewhat tentatively) be arguing in the third section of this chapter.
Of course, one is left to wonder about this “I” who recognizes and autonomously adopts reasons prior to the existence of me. \(^{85}\) In short, on either reading of the relationship between agents and maxims, we are left with a kind of circularity that will be drawn out below as the *identity/reasons problem*: Does the agent make the maxim/logos, or does the maxim/logos make the agent? If Korsgaard continues heading in the latter direction, her view might seem to fit better with the metaphysics of the Modern Scientific World View with respect to the origin of “agents” because she would not be simply starting out with free agents accoutered with “reasons.” This would seem to be the path of Gibbard or Post or, perhaps more consistently, Michael Ruse. \(^{86}\) But it has been shown in Chapter 3 (and will be argued also in connection with John Post in Chapter 5) just how difficult it is to understand how there could be *reasons* at all in such a view. *Causes* there are aplenty, but *reasons* go missing; that is, unless one is willing to get metaphysical in a way that seems highly objectionable from within the metaphysics of the Modern Scientific World View and start speaking of the *logos* in a very robust way indeed, such that reasons just are there somehow for us. This way lie Moore or Plato—scarcely views that sit easily within a materialist metaphysic (as J.L. Mackie made clear). Prioritizing reasons over identities will create its own problems for materialists, like Korsgaard, as shall be argued below. Where will “reasons” reside without persons, and what structure could itself *be* a reason? But without reasons guiding or shaping identities, nothing remains to rule out objectionable identities and establish a reflective humanity that strives toward the unconditioned as *the* human identity. Either way, it seems very difficult to come up with a starting point for the whole

---

\(^{85}\) Here the identity/reasons fault-line running through Korsgaard’s work is merely pointed out (see also note 71 above). This fault-line and its significance will be explored in section below.

\(^{86}\) Ruse, “Social Darwinism Updated?”
business of there being reasons at all for moral philosophers attempting to work within the metaphysics of the Modern Scientific World View that Korsgaard wants to be true.

In summary, then, there are two possibilities open to Korsgaard here at this pivotal point in her moral philosophy regarding the relationship between rational agents and maxims. On the one hand there is the identity-priority maxims possibility in which maxims are candidate-reasons that are transformed into reasons by the reflective endorsement of a human agent. This gives us the moral law because all of us are committed to the reflective identity of a human (that is, a rational agent capable of conferring goodness on ends), this being necessary to have any practical identity at all. As Michael Smith summarizes Korsgaard’s point,

when I look for a description of the cares and concerns of my reflective self, I cannot doubt that it would be appropriate to describe my reflective self as having the cares and concerns of a creature who is capable of forming desires as the result of reflection. It is, after all, in the nature of my reflective self as a reflective self to have such cares and concerns. It therefore follows that I cannot question this conception of my practical identity. It makes no sense at all to ask whether I should have this practical identity, because I cannot have any practical identity at all without having this one.87

Any action arises from some practical identity or another, any one of which commits us to valuing humanity, and this not only in ourselves, but in any being who possesses the relevant capacity for reflective endorsement. “If we regard our actions as rational, we must regard our ends as good; if so, we accord to ourselves a power of conferring goodness on the objects of our choice, and we must accord the same power—and so the same intrinsic worth—to others.”88 Reasons emerge from the practical identity of agents, and because all

87 Michael Smith, “Critical Study: Search for the Source,” The Philosophical Quarterly 49, no. 196 (July 1999): 390-91. See also pp. 272-3 above. It should be noted that Smith is critical of Korsgaard’s view.

88 Korsgaard, Creating the Kingdom of Ends, 261-2.
agents share in the identity of the human, all agents share some of the same basic reasons
flowing from that identity.\textsuperscript{89}

On the other hand one might take maxims to be \textit{form-priority maxims} in which
maxims are reasons by their very structure of relating an act to an end, and that any time an
agent has a maxim it is already a reason in virtue of its form. There are well-formed maxims
and ill-formed maxims, and the latter are not maxims at all. There are maxims and
schmaxims. Any maxim is a reason in virtue of its form, a form bound up with the very
existence of humans themselves. In the strongest sense, we have to will according to this
form—the form of a maxim—to be human at all.\textsuperscript{90} With a true maxim, its reflective
endorsement is already included in its formulation, \textit{with the form drawing out the will}. \textit{There is no
freedom apart from this form of relating act and end that is itself constitutive of willing.}

Schmaxims, however, are never reasons, because of a failure of their form. They are
ersatz maxims where an ununiversalizable inclination has corrupted what superficially
resembles an authentic instantiation of a maxim; the inclination dominates the behavior, so
that the action fails to arise from the form of the maxim itself. This may be related to a
distinction Korsgaard draws between an action and a mere act, where only the former is tied
to a maxim, where that which is chosen “is the act for the sake of a certain end, and it is the

\textsuperscript{89} It is not clear how far this goes toward answering the challenge Raymond Geuss puts forward so
forcefully in the closing pages of his critique of Korsgaard, “Morality and Identity,” in Korsgaard, \textit{The Sources of
Normativity}, 197-99. His critique is that while (at best) Korsgaard may have shown that each of us must value
our own humanity, she has not shown why each of us must value the other’s humanity. See the passage at
note 46 above.

\textsuperscript{90} It is not clear exactly how either the maxim-form or the human identity are supposed to fit into a
materialist view of reality. In an understated way Gibbard points this out when he is critical of Korsgaard’s
“elaborate Kantian package, thinking that their similar starting points afford only “something more modest,”
(Gibbard, “Morality as Consistency in Living,” 141). Such tensions will be explored briefly below.
whole thing—this act for the sake of that end—that is the object of choice.” When the lawlike form is not present in the formation of the act-end pair, so that it is the content of an inclination that dominates the pairing and any consequent “choice” of the human organism to perform this mere act—when, that is, a maxim is not present—then what has gone missing is precisely choice itself. However, when the lawlike form is present, so that there is a maxim, then the will that is free determines itself by following this form rather than the inclinations embedded in the myriad schmaxims that appeal to us on the level of mere, non-rational animality: “suppose the will chooses the maxim of self-love. In that case, it departs from its position of spontaneity and puts itself in the service of inclination.” Only in the form of thought that doubles back on itself and makes its own internal consistency its object do we have the freedom of thought from the external domination of inclination.

On this reading, whenever a maxim brings to expression the lawlike form of truly free thought, there is a reason there—whether the human acts as a human and chooses the one thing that allows her action to be free, or instead collapses back into the domination of inclinations. As Korsgaard puts it, “a maxim is an entity whose intrinsic properties make it a final reason for action, a final ‘right’. Something which has the form of a law, that is, which is a law by virtue of its internal structure, is intrinsically suited to answer the question why the action it dictates is necessary. In this sense, a good maxim is exactly the sort of entity which the realist argument requires.” Reasons, then, would be external in the sense of being tied to this lawlike form of reasoning that comes to expression in the maxim, a form that is what

---

91 Korsgaard, “Internalism and the Sources of Normativity,” 53.

92 Korsgaard, Creating the Kingdom of Ends, 166.

93 Korsgaard, The Sources of Normativity, 112; emphasis in original. Korsgaard continues on to emphasize that the form of realism she is advancing is “procedural rather than substantive realism,” (112).
it is independently of anyone actually willing in accordance with it and so being a free, reasoning being. On the *form-priority maxim* strand of Korsgaard’s moral philosophy, morality just is freedom, which is in turn the maxim-structure of thought.

It seems likely that the identity-priority maxim strand is Korsgaard’s official stance (or at least was at the time of her fullest statement of her moral philosophy in *The Sources of Normativity*), but there is a strong undercurrent of tension throughout her work that pushes her toward the form-priority maxim strand—one that has dramatically increased in her most recent writings. This tension is endemic to the kind of position she wishes to defend. Korsgaard is trapped in a circle—the identity/reasons problem—where taking either stance lands one in difficulties that leave no way forward. The only solutions are by metaphysical fiat, and in both cases—whether one mysteriously grants oneself the human agent or the maxim/logos—the resulting stance does not square with the metaphysics of the Modern Scientific World View. In other words, it is not clear how Korsgaard’s conception of the human agent and its first-person standpoint is supposed to fit with her materialism (a problem that will be explored in greater depth in the third section of this chapter). In the identity-priority maxims view, Korsgaard just *starts out* with a fully accoutered human self, with all its ability to appreciate, weigh, and act on normative reasons.94 This is a *lot* for a materialist to grant herself; an autonomous moral agent just appears on the scene. But it is not clear that the maxim-structure essential to the form-priority maxims view—a structure somehow existing independently of and calling out the human agent—would be any more

---

94 Showing how there could be such a human agent in a materialist world is basically what Heidegger spent his entire career trying to answer. Korsgaard is right to take the human agent seriously and to start moral enquiry there; what is left unexplained is how it can be right for a materialist to do so. As Paul Moser commented (personal communication, 6 June 2008), Korsgaard seems to be “using the originally biological category of ‘human’ in a way that abandons materialism.”
materialistically kosher. Either way, Korsgaard starts out assuming something of dubious materialist credentials.

In the next section it will be argued that whether Korsgaard takes the identity-priority maxims view (as seems to be her official stance) or the form-priority maxims view, she faces severe difficulties. The second section will press the identity/reasons problem against Korsgaard’s moral theory, exposing her inadequate account of moral reasons and arguing that she fails to avoid moral reasons arbitrariness. After that, the third section will suggest that Korsgaard’s thought suffers from a “transition problem” much like the one Kant attempted to answer in the *Critique of Judgment*, and that if there is such a connection it makes clear that the identity/reasons problem is a fundamental difficulty for Korsgaard, one that might be of wider import to contemporary Kantian moral philosophies more generally.

II. Identity Problems

A. The Identity/Reasons Problem

Korsgaard’s position that a person’s moral reasons flow out of an identity that the person endorses creates a fundamental difficulty for her. Put crudely, there is a chicken-or-the-egg problem involving reasons and identities: Does a person’s identity—any identity whatsoever—lead to the moral reasons one has? Or does the maxim-structure itself provide reasons that somehow lead to the identity one has, those reasons somehow existing prior to any practical identity that endorses them? If the former—the identity-priority maxims

---

95 In her latest work Korsgaard calls what I call the “identity/reasons problem” the “paradox of self-constitution,” (Korsgaard, *Self-constitution*, 20, 41). I will retain my terminology, which, in any case, gets at the heart of the “paradox.”
account—then one’s identity comes prior to reasons for endorsing anything, including the identity itself. The endorsement of an identity responds neither to that identity nor to reasons, and hence is arbitrary, influenced by factors adventitious from the moral standpoint itself. On the other hand, if Korsgaard were to pursue the form-priority maxims account, then the reasons emerging from the very form of the maxim somehow preexist the endorsement or acceptance that leads one to choose one identity over another. In this case, since a maxim is an act-end pairing, an end would have to preexist any particular endorsing agent—these maxims themselves calling out and constituting such agents. It is, of course, very difficult for a materialist to sustain such a position. Nor is it clear why just the maxims enshrining the ends Korsgaard would like to see elevated above other ends are the ones that would emerge in the primordial constitution of agents, nor why it should be the case that all agents should then apply these ends beyond the circle of their own particular agency. In short, Korsgaard seems to be caught in a deeply problematic circle: on the one hand, the practical identities that give rise to reasons cannot themselves be chosen for reasons, with the result that moral reasons are arbitrary due to the arbitrary nature of the identities that set them; on the other hand, however, the maxim-structure fails to set non-arbitrary moral reasons apart from a rational identity able to endorse only maxims that are “good,” with the result that moral reasons arbitrariness infects the view. That is to say, a maxim must be tested to see if it is a law, and this must be done by an agent who brings a demand for consistency and universality to the maxim.⁹⁶ So, the second tack also faces a severe difficulty: the maxim

⁹⁶ As Korsgaard puts it, in a very important passage, nothing determines what the law must be. All that it has to be is a law. Therefore the categorical imperative is the law of a free will. It does not impose any external constraint on the free will’s activities, but simply arises from the nature of the will. It
form cannot set moral reasons apart from the agent. Moreover, to add insult to injury, it’s not clear how either tack will be made to fit with materialism.

In either case, whether Korsgaard takes identities or reasons as primary, it turns out that reasons and identities relevant to human moral life owe their existence to non-moral contingencies (the evolutionary history that shapes the identities that endorse one thing or another, or whatever it might be that could give rise to the appropriate maxims) that serve as the brute givens of the moral life. This is essentially the “substantivalist” project that Allan Gibbard urges in place of Korsgaard’s “proceduralist” project. Calling his view a form of “substantivalism,” Gibbard claims that the only “why stoppers” available to us are “those substantive principles that seem most evident, those policies for living that we find most clearly unproblematic” because of our evolutionary history. In other words, Gibbard sees that Korsgaard’s “proceduralism” moves in a circle and cannot stop the “why” questions, which must rather come to an end in a substantive fact that has explanatory force and gives us all that is meaningful in our demand for reasons, for “why?”’s. This seems right, but, when these substantive principles find nothing to draw upon except the forces operative in Gibbard’s “Galilean” metaphysics, they fail to underwrite reasons that address moral agents. In essence, Gibbard finds a grounding for morality, but one that can’t underwrite moral reasons that address moral agents with authoritative claims, while Korsgaard has moral

describes what a free will must do in order to be what it is. It must choose a maxim it can regard as a law.” (Korsgaard, The Sources of Normativity, 98; italicized emphasis in original, underlined emphasis added).

Clearly the spontaneous will—particularly in legislating consistency over itself—precedes and ratifies some maxims as laws, insofar as these particular maxims meet with the demand for consistency that is brought to the maxim due to the nature of the will that is brought to the maxims. There are, as Korsgaard puts it, “good” maxims—implying that there are “bad” maxims—with the rational agent preceding them and pressing a test upon them based on its own rational nature and the consequent demand for exceptionless consistency.

Allan Gibbard, “Morality as Consistency in Living,” 142.
claims addressed to agents, but can find no grounds for them—since the “Galilean”
metaphysic she too accepts can provide none and the procedures she finds move in a
heautonomous circle (as shall be argued in the third section, below). Chapter 3 already
exposed the flaws of Gibbard’s attempt to pursue a “substantivalist” tack as a materialist.

But if Gibbard is right, Korsgaard’s putatively “proceduralist” project turns out to be a
substantivalist project after all, with the first personal standpoint of making a decision finally
needing roots in the non-personal stuff that gives rise to everything else within the
materialist’s world.

If Gibbard’s way forward is untenable, what can be said for Korsgaard’s moral
toory theory when faced with identity/reasons problem? Can she extricate her theory from the
identity/reasons circle? On balance, Korsgaard seems to lean toward trying to show how an
identity can proceed reasons and yet avoid moral reasons arbitrariness, although it appears
that her latest work may be starting to lean toward seeing the structure of maxims (logos)
preceding and somehow calling forth the identity. Thus, in her very recent book, Self-
Constitution, we find Korsgaard admitting that “this is where things get complicated,” and
going on to say that while one might take the “prior unity of the agent” to be essential for
genuine action on the part of that agent, this “cannot be how it works.”

Instead, “in the
relevant sense there is no you prior to your choices and actions, because your identity is in a
quite literal way constituted by your choices and actions.” A good action, then, “is one that

98 Korsgaard, Self-Constitution, 19.

99 Korsgaard, Self-Constitution, 19; emphasis in original.
both achieves and springs from the integrity of the person who performs it.” This is a dramatic statement of what I am calling the “form-priority maxims” view.

Korsgaard has painted herself into a corner I have called the identity/reasons problem. In her most recent work she has seen the difficulty, calling it the “paradox of self-constitution”:

It is as the possessor of personal or practical identity that you are the author of your actions, and responsible for them. And yet at the same time it is in choosing your actions that you create that identity. What this means is that you constitute yourself as the author of your actions in the very act of choosing them. I am fully aware that this sounds paradoxical. How can you constitute yourself, create yourself, unless you are already there?

As I have put it, does identity precede reasons in the maxim-structure of thought, or do reasons precede identity? This is indeed a paradox. In what follows, both of these ways of trying to deal with the identity/reasons problem will be shown to be inadequate, both of them finally leading into the pit of moral reasons arbitrariness.

---

100 Korsgaard, Self-Constitution, 25. Again, there is a very close parallel to a problem that greatly exercises Heidegger, one that Heidegger perhaps has seen more clearly than Korsgaard has and one that arises for each of them for related reasons. Perhaps the problem can be seen most clearly in a passage from William J. Richardson’s insightful reading of Heidegger’s Kant and the Problem of Metaphysics: “Through this function of donation to self there is fashioned by the pure intuition a view (Anblick) of either space or time. This view is received by the intuition itself, but the reception is the very thing that constitutes the donation. Hence the pure intuition is an affecting of itself, sc. self-affection,” (William J. Richardson, Heidegger: Through Phenomenology to Thought, 4th ed. (New York: Fordham University Press, 2003), 116). This is more or less just the project Korsgaard has begun to pursue; it is a question of the relation of spontaneity and receptivity in a world where there is no pre-existent Logos. Obviously this matter cannot be pursued in this dissertation, but it is helpful to get a glimpse of the more fundamental problem that Korsgaard is running into. Korsgaard, too, needs some sort of self-affection in which the subject gives the reasons it receives or the reasons give the subject that receives them in some sort of coeval primordial irruption of the cons of non-personal stuff. Heidegger spent his life wrestling with this one problem, and, one thinks it safe to say, never really got beyond asking the question. His intensity in asking this question is what I find to be of most value in Heidegger’s work.

B. Setting Ends: Identity before Reasons

Korsgaard knows very well the problem facing her moral philosophy when she places identity prior to reasons as the originating point of morality. Since nothing in the materialist world of a Galilean metaphysic is of the right type to address reasons to agents who must decide what to do, moral reasons must originate with contingent rational beings such as humans (of course, this also creates the severe problems for the form-priority maxims view that will be seen below, and thus also serves as the impetus for the identity-priority maxims view). Korsgaard tries to ground moral reasons in the fact that we are beings who must decide what to do by taking up a practical identity. The question is how to move from prudential reasons tied to one identity or another to what Korsgaard calls “justifying reasons” that can adjudicate between the conflict of reasons arising from the myriad possible practical identities and the reasons flowing out of them. Korsgaard wants to avoid the moral reasons arbitrariness that arises if moral reasons are merely prudential reasons tied to wildly varying practical identities. But if such identities precede reasons, then it would seem that the reason-setting identity could take a wide variety of shapes, with the result that moral reasons would vary wildly depending on the motley identities that give rise to them (or, alternatively, that they don’t vary all that much because of the non-moral forces that grip those identities—as Gibbard would have it). As Korsgaard notes, the “rational will” she sees as fundamental to humans as such, “must have a principle. Because it is free, it must choose this principle for itself. Nothing determines this choice: it is completely spontaneous. Since its principle determines what it counts as a reason, nothing yet counts as a reason for it. But if nothing yet counts as a reason for it, it appears to have no basis for
choosing its principle.”

Korsgaard immediately moves on to her proposed solution to the problem, a solution one finds her reformulating over and again in various writings. But before delving into her attempt to solve the problem, it will be well to sharpen the problem a little more.

Recall that Korsgaard follows Kant’s lead in holding that the “capacity to propose an end to oneself is the characteristic of humanity (as distinguished from animality).” In summing up this Kantian idea, Korsgaard says that “ends are ‘set’ by practical reason.” Elsewhere Korsgaard states that human beings “confer value” and “bring goodness into the world,” and that activities that have value “must actually get their value from our valuing them.” “The good,” Korsgaard holds, “cannot be contemplated but only created by our efforts. What initially looks like a sort of moralism on Kant’s part is really the consequence of his humanism. The only value there is is that which human beings give to their own lives. We must be the source of value.”

102 Korsgaard, *Creating the Kingdom of Ends*, 25. Even if this move were to succeed, it’s unclear what status the “reason” thus produced by the subject for the subject would have. This problem will be explored below in section three as a question of what Kant called “heautonomy” in his own attempts to solve a “transition problem” in the architectonic of his critical philosophy.


104 Korsgaard, *Creating the Kingdom of Ends*, 110. Other places where Korsgaard speaks of “setting ends” include Korsgaard, *Creating the Kingdom of Ends*, 22, 114, and 260. Korsgaard speaks of humanity as the “source of value” on 144 and 246.

105 Korsgaard, *Creating the Kingdom of Ends*, ix (this language also appears in 122, 240-41, and 260-61, among other places).

106 Korsgaard, *Creating the Kingdom of Ends*, 273.

107 Korsgaard, *Creating the Kingdom of Ends*, 246.

108 Korsgaard, *Creating the Kingdom of Ends*, 246.
the world somehow,” and, in a practical sense, human beings are the ones who must do it. Korsgaard makes it plain that moral value does not preexist the activity of human beings by which value is first created. As Korsgaard sees it, it is vitally important that we follow Kant in recognizing “the role of the good will in conferring value upon the ends of the person who has it.” Human “reflective endorsement” confers value, sets ends, and is the source of goodness—and moral reasons do not precede this human activity. The human activity of reflective endorsement first brings moral reasons into the world. As Korsgaard puts it, “the test for determining whether an impulse is a reason is whether we can will acting on the impulse as a law. So the test is a test of endorsement.” Reflective endorsement precedes reasons.

What, then, guides reflective endorsement? In The Sources of Normativity Korsgaard makes it clear that a person’s practical identities set what will be reflectively endorsed by that person. In the final analysis, as was seen above, “the identities give rise to reasons and obligations. Your reasons express your identity, your nature; your obligations spring from

---

109 Korsgaard, The Sources of Normativity, 5.

110 If true, of course, this would rule out the form-priority maxims account. I believe that Korsgaard thinks this is the case for value generally. Indeed, if it does nothing else, Mackie’s “queerness” argument captures the attitude of many materialists who cannot imagine what a value could be in a world of only natural stuff. Values emerge along with human beings somehow, whether through expressing some plan (Gibbard), through a Kantian reflectivity characterizing human practical identities (Korsgaard), through what we have been selected for (Post), or in some other way connected with human beings arising in the natural world.

111 Korsgaard, Creating the Kingdom of Ends, 260. On the following page, Korsgaard states: “the Kantian approach frees us from assessing the rationality of a choice by means of the apparently ontological task of assessing the thing chosen: we do not need to identify especially rational ends. Instead, it is the reasoning that goes into the choice itself—the procedures of full justification—that determines the rationality of the choice and so certifies the goodness of the object. Thus the goodness of rationally chosen ends is a matter of the demands of practical reason rather than a matter of ontology,” (261).

112 Korsgaard, The Sources of Normativity, 108; emphasis in original.
what that identity forbids.” Or, again, in her earlier work Korsgaard urges her readers to see that the will has “a principle from which it will derive its reasons. The principle it chooses will determine what it counts as a reason. But precisely because at this ‘moment’ the will has not yet determined what it will count as a reason, it seems as if there could be no reason for it to choose one principle rather than another.” It would seem, then, that there are no reasons that guide the choice of an identity or a principle, for it is a human activity of endorsement that gives rise to reasons in the first place. But this would mean that the choice of an identity or a principle is for no reason, and the arbitrariness of this choice leads to an arbitrariness infecting moral reasons generally. Identities set reasons, leaving no reasons to adopt one identity over another. The conflicting reasons flowing out of the various practical identities each share the same grounding in a practical identity someone has taken up. It would seem that Korsgaard has ended up in the “coal pit” of moral reasons arbitrariness.

Korsgaard, however, thinks she has a way of avoiding the path down into the pit. The key is that there is a certain identity that is both unavoidable and that constrains itself to move in one direction over another—and Korsgaard argues that this very self-constraint constitutes morality. As Korsgaard sees matters, if there is no human reflective identity performatively necessary to taking up any practical identity, then we will be in grave danger of moral reasons arbitrariness (or, in her terminology, a paralyzing inability to justify one moral action over another in our own eyes). So, while reasons do indeed flow out of a human activity of endorsement (any endorsement, as Korsgaard makes clear in interacting


114 Korsgaard, Creating the Kingdom of Ends, 165.
with the Mafioso problem, this activity itself is constrained by an identity that reflectively turns in on itself and in this reflection non-arbitrarily gives rise to reasons that it gives to itself. This unavoidable, self-reflexive identity is what Korsgaard calls “humanity.” “Since you cannot act without reasons and your humanity is the source of your reasons you must value your own humanity if you are to act at all.” This, however, just is to value self-reflexivity as the capacity to have reasons. There is a performative necessity in acting according to any practical identity that pushes us to value the very capacity of rational choice that is of the essence of a human identity. Anyone who acts at all cannot dismiss the value of reasons without falling into performative contradiction, and so beings who act cannot help but value humanity—in themselves and, Korsgaard will argue, in others. As Korsgaard puts it, “we must regard others as capable of conferring value by reason of their rational choices and so also as ends in themselves.” The fundamental demand we make on ourselves in the very nature of our act of deciding what to do is that we act for a reason rather than being swept along by one force or another that would determine our behavior; but acting for a reason, again, just is the determination to decide for oneself rather than being swept

---

115 See pp. 271-2 above.

116 As pointed out above (pp. 291-94 above), Korsgaard understands “humanity” in a specific, Kantian sense. While Korsgaard clearly takes animals into account in her moral philosophy, she also emphasizes the unique capacities of human beings with respect to appreciating, weighing, and acting on reasons (see note 81 above).

The self-reflexivity at the heart of Korsgaard’s account points toward the heautonomy at the heart of Korsgaard’s attempt to answer what Kant called the transition problem that beset his own philosophy and that dogs Korsgaard in turn. It is this problem that Heidegger, too, sought to solve by postulating a sort of self-affection or self-donation (see notes 100 and 22 above). These issues will be explored in the next section of this chapter.

117 Korsgaard, The Sources of Normativity, 123.

118 Korsgaard, Creating the Kingdom of Ends, 260.
along by one force or another, and so involves giving oneself one’s own law. It is the determination to act freely, to act according to a law one gives oneself. “At the standpoint of spontaneity, the will must, in order so to speak to commence operations, choose a principle or a law for itself. Nothing provides any content for that law. All that it has to be is a law.”¹¹⁹ It is the very form of making a decision that matters here, a form that dictates a consistency in the decision that one makes.¹²⁰ To do otherwise is to attempt to make a free choice not to choose freely—the very attempt being incoherent and self-stultifying. One attempts to will what cannot possibly be willed. Moral reasons arbitrariness simply cannot be a problem for the human being as such, since the very nature of humanity is to have moral reasons that are not arbitrary. One cannot be a human being at all unless one has non-arbitrary moral reasons, that is, unless one freely decides what to do. Humanity—all humanity, wherever it is found—must be valued, if the agent is to make any choice at all rather than being swept along “pathologically” by inclination.

But Korsgaard’s arguments here fail to work in their own terms. The first step in making plain the problem Korsgaard faces is to return to the problem of the contented criminal from Chapter 2. This is a problem similar to the “idealized Mafioso” problem G.A. Cohen initially posed to Korsgaard in his comment on her Tanner Lectures that would become The Sources of Normativity. As we saw above, Cohen’s “idealized Mafioso” has a practical identity that leads him to do all sorts of hideous things, his reasons for those actions owing to what this identity leads him to endorse. This practical identity would seem

¹¹⁹ Korsgaard, Creating the Kingdom of Ends, 166; emphasis in original. One can see the germ of the form-priority maxims strand of Korsgaard’s thought in this passage. The next section explores the reasons for this fault-line running throughout Korsgaard’s work.

¹²⁰ One sees here the impulse in Korsgaard’s thought that pushes toward the primacy of the structure of the maxim (logos) over the human identity. This alternative story will be critiqued in the next section.
to underwrite reasons and obligations as much as any other identity, and the individual who holds this identity seems free to do so: “he can prescribe the Mafia ethic to himself.”\textsuperscript{121} And in doing so, he would seem to have ground as good as anyone else for doing so. He certainly has \textit{endorsed} the identity, and endorsement is that which gives us reasons and obligations. As was seen above, Korsgaard insists that “there is a sense in which these obligations [arising from the Mafioso practical identity] are real—not just psychologically but normatively.”\textsuperscript{122}

But has he \textit{reflectively} endorsed the identity? This is crucial, for there is another sense in which bare endorsement does \textit{not} afford us reasons. For this more robust sense of reasons reflection must be added. “Normative reasons,” in Korsgaard’s view, “always come from reflective endorsement.”\textsuperscript{123} So, what is it that the “reflective” part adds to bare endorsement that transforms it into something that gives human beings reasons? Korsgaard has something more than mere subjective consistency in view, something one person could have within one practical identity that could be in some sort of conflict with the consistent view of someone else within a different practical identity (for example, the Mafioso and Elliott Ness). Korsgaard has in mind something that cuts across such conflicts and at least in some cases uniquely justifies one course of action over another. But it’s hard to make out just what she could have in mind here, because for Korsgaard—as for Gibbard as well—it really is \textit{bare human endorsement} that does the true normative work. A person is committed to acting in a particular way because he has endorsed a principle or an identity that leads to the


\textsuperscript{122} Korsgaard, \textit{The Sources of Normativity}, 257.

\textsuperscript{123} Korsgaard, \textit{The Sources of Normativity}, 252.
doing of that action. The “to be done-ness” of a course of action has that feature because someone endorses it; it’s not as if it wears it on its sleeve, by some sort of magic (given the Modern Scientific World View, that would be “queer”). Endorsement by someone or other is the origin of normativity. This becomes very clear when Korsgaard bites the bullet in response to Cohen’s Mafioso problem. She recognizes that it would be “intellectually tidy” and “spare [her] trouble from critics” if she could say that the Mafioso does not actually have any obligations or reasons flowing out of his endorsed identity. In other words, she recognizes that it would be nice if she could say that the Mafioso’s endorsement of his practical identity gives him “no normativity, only the psychic appearance of it.” But Korsgaard flatly rejects this. Indeed, this is what she cannot do, for it really is bare endorsement that does the true normative work. Human beings have reasons only as they endorse something—it is just this that brings reasons into the etiolated world of the Modern Scientific World View. No endorsement, no reasons. And so Korsgaard bites the bullet on the Mafioso problem and allows that there is a sense in which the reasons he has are normatively real. Normativity arises not merely from a practical identity that is reflectively endorsed, but also from an identity that is non-reflectively endorsed. In its originating impulse, then, normativity is extremely pliable, arising wherever a human being endorses a practical identity. Reasons and obligations follow endorsement.

The trick for Korsgaard (or one of the tricks, at any rate) is to show how not just any endorsement is as good as any other. Gibbard’s more obviously evolutionary account fails in this regard, as I have argued. But Korsgaard pins her hopes to reflectivity. The question then presses all the more forcefully: What does reflection add to this picture?

---

124 Korsgaard, The Sources of Normativity, 257.
Strictly speaking, normativity doesn’t need reflection, only endorsement. And yet Korsgaard still holds that the Mafioso has reasons and obligations to give up this immoral practical identity. She rejects the idea that “the Mafioso’s obligation to give up his immoral role is something that exists only from the perspective of the rest of us, and not in his own. For he is a human being, who arrives at his reasons through reflection. And the activity of reflection has rules of its own.”125 Apparently, what Korsgaard needs is some sort of performatively necessary endorsement that no one can avoid, and she thinks reflection can provide what is needed. It turns out then, that not all obligation and not every reason arises from conscious endorsement, for the Mafioso has reasons and obligations arising from an identity that he does not endorse—or at least that he is not aware of having endorsed. But the very adoption of any practical identity whatsoever renders performatively necessary (if not consciously acknowledged) his endorsement of the value of his own first-personal ability to choose, the lawful form necessary to all choice whatsoever (the alternative being domination by inclination and thus the forfeiture of spontaneous human willing), and the prescription of his choices and their form freely to himself as an unavoidable aspect of having to act within a first-person standpoint he cannot avoid.126 This of course cuts the nerve of the motivational internalism that is part of the attraction of making endorsement central to one’s account of morality. But the main questions lie elsewhere.

Korsgaard clearly makes morality depend on a notion of humanity that (1) is practically unavoidable for human beings, and (2) is characterized by reflection that demands consistency in what one wills, where this is sufficient to ground human morality.

125 Korsgaard, The Sources of Normativity, 257; emphasis in original.

126 There are echoes of Sartre’s thought about “bad faith” here. Performatively, we cannot foist our choices onto something that excuses us by determining our choices.
Korsgaard’s moral philosophy in *The Sources of Normativity* finally appeals to the same pattern of thought that characterized her earlier work in the various essays collected in *Creating the Kingdom of Ends*: “All that it has to be is a law.” In the end, Korsgaard’s appeals to identity and endorsement lead back to a notion of humanity conceived in a Kantian manner—a notion of humanity that cannot be avoided on pain of performative contradiction (is a “practical requirement”) and where consistency underwrites morality. “You make yourself an end for others; you make yourself a law to them. But if you are a law to others in so far as you are just human, just *someone*, then the humanity of others is also a law to you. *By making you think these thoughts*, I force you to acknowledge the value of *my* humanity, and I obligate you to act in a way that respects it.”\(^{127}\) The identity that matters is a human identity, where this is conceived in Kantian terms: “our identity as moral beings—as people who value themselves as human beings—stands behind our more particular practical identities.”\(^{128}\)

But this creates a host of difficulties. As already appeared above, Allan Gibbard, for one, is quite skeptical that consistency can get the job done apart from some kind of raw input from outside the system. As Gibbard puts it, “our judgments of how to live depend not just on a logic of what to do, but on our proclivities to ‘see’ some eventualities as goals to pursue, and others as dangers to shun. If our proclivities go wrong, then we go wrong in our life policies, be they ever so consistent. Illogic isn’t the only way to be wrong.”\(^{129}\) One

\(^{127}\) Korsgaard, *The Sources of Normativity*, 143; emphasis in original. In summing up his objections to the line of thought Korsgaard advances here, Michael Smith states that “it is simply wrong to suppose, as Korsgaard does, that I must value leading a life like the life that my reflective self leads,” (Smith, “Search for the Source,” 394).

\(^{128}\) Korsgaard, *The Sources of Normativity*, 121.

\(^{129}\) Gibbard, “Morality as Consistency in Living,” 146.
is reminded of the possibility Gibbard raises of an ideally coherent anorexic or Caligula. Mere consistency in willing does not clearly give us non-arbitrary moral reasons without the right kind of external input—at best it gives us internally consistent subjective willing (that may well conflict with the internally consistent subjective willing of others). Korsgaard (rightly) wants something more robust than this.

There are other problems for mere consistency as the ground for non-arbitrary moral reasons. If indeed you must value your own humanity (it’s not clear that Korsgaard has even established this, but let that pass for the moment), that doesn’t mean you must value the putative humanity of someone else. The value of that other person’s humanity must be established somehow—consistency alone fails to do this. Perhaps you must value yourself from within your own first-person standpoint, but that itself does not determine how other people are to be viewed within that first-person standpoint. Nothing within that standpoint dictates that those other primates outside oneself be viewed in that first-person standpoint as similar to you—the one who has this particular viewpoint and is making this particular decision. Nothing about this requires viewing such animals as factoring into the decision in a particular way. Perhaps they are Jews, or women, or blacks, or Sudanese Christians, or whatever—and nothing about your first person perspective requires taking those kinds as relevantly similar to yourself. Indeed, those kinds could enter into someone’s decisions not necessarily as a person, but as something fully explained in third-person, materialist event-causal terms.\(^\text{130}\) The explanation one accepts about these particular animals can indeed shape how they are viewed in the first-person standpoint. For Mao, the peasants he slaughtered were “two shoulders and bum,” and it’s doubtful that mere inconsistency

\(^\text{130}\) See Chapter 2, section I.B on the distinction between reasons and causes.
was the problem. Rather, he held beliefs that undermined the place of other human beings as he decided what to do in the first-person standpoint, so that he could take himself to be being consistent while ignoring their protests.\textsuperscript{131} He held beliefs such that he could take himself to be consistent from the viewpoint of that belief system. The point is that in deciding what to do there are multiple different ways humans can be viewed within the standpoint from which a decision must be made. The basic problem facing Korsgaard’s version of a universalizability argument is that others need not be brought into the first-person perspective in the same way that might performatively necessitate the valuation of oneself. \textit{Consistency} alone cannot determine this. And appeals to awareness of the facts won’t help here. Indeed, the materialist’s facts may in fact push one toward \textit{not} assigning objective value to oneself either\textsuperscript{132}; it’s just that \textit{performatively} you can’t help but do so. But here’s the point again: \textit{You are in no such performative stance with anyone but yourself.}

\textsuperscript{131} As Paul K. Moser has pointed out (personal communication, 6 June 2008), “even if his beliefs were evidentially arbitrary at points, they could be consistent.” Indeed, I wonder whether or not Mao \textit{might possibly}, as an adherent of the Modern Scientific World View, have been able to argue that his beliefs were not even \textit{evidentially} arbitrary, in that he might have acknowledged that they shared the same biological status as himself, thus avoiding evidential arbitrariness, and yet did not value that in them: “Yes, they are homo sapiens and they have the same capacities for thought and feeling I have, but I don’t see a reason to value that \textit{in them}. I’m just happy I was clever enough to avoid being in their position. Indeed, it’s good for the future of humanity that the clever survive.” I find it difficult to see even an evidential arbitrariness there (\textit{though I’m not committed to defending that stance here}), given the World View he accepts. Here I throw my hat in with Gibbard; consistency alone just won’t get moral philosophy where it needs to go, and piling on facts won’t necessarily help matters. For all that I do think that Mao got something \textit{wrong} and was in a position to know it at some level, but that is \textit{not} the subject of this dissertation.

Finally, even if it could be shown that Mao was violating a value of consistency and that this value holds within the Modern Scientific World View, I think someone who accepts the Modern Scientific World View would find it difficult to craft a good argument that this value \textit{should} be valued more highly than all other values that might somehow also be identified within that view of reality (e.g., survival).

\textsuperscript{132} Indeed, as Nagel admits in his own version of a universalizability argument in \textit{The Last Word}, the view that one is objectively worthless (a view I think highly plausible, given a materialist metaphysic) “satisfies the generality condition for reasons” and is “perfectly consistent” (122). It’s important to note how Nagel resists embracing the view: It is, Nagel states, “in my opinion highly unreasonable and difficult to accept. Can you really believe that objectively, it doesn’t matter whether you die of thirst or not?” (122). This, of course, is to appeal to \textit{performat}ive factors. \textit{I} can’t do that for \textit{myself} because, well, I’m \textit{me}. And I may realize that \textit{you} can’t do that for \textit{yourself} because you’re \textit{you}. But this performative factor drops out when I think \textit{of you} for the simple
The insufficiency of consistency, however, is not the only problem here. Consider the contented criminal, who is much like Cohen’s idealized Mafioso, except this criminal is known to be content in her practical identity and the life that it has given rise to. She—like Mao—has rightly calculated that this life is viable, that her identity can be sustained long term in the face of any resistance from others, and she is content in the life that she leads. Now Korsgaard would enjoin her to sustain her reflection until she finally comes to see that she is obligated to give up the viable criminal life with which she is contented. One of the rules of reflection incumbent on all humans, “and perhaps the most essential,” Korsgaard informs her, “is the rule that we should never stop reflecting until we have reached a satisfactory answer, one that admits of no further questioning.” Now it is quite likely that this will have no traction whatsoever with our contented criminal, but leave that aside. It’s not likely that many moral views would change the mind of the contented criminal. There

---

133 The concept of viability appears in the work of Richard Brandt, and was explored above.

134 Korsgaard, The Sources of Normativity, 258. In personal communication (6 June 2008) Paul K. Moser has drawn attention to a certain similarity between Korsgaard’s thought and that of Alan Gewirth. This similarity can be readily seen in a summary Gilbert Harman gives:

According to Gewirth, you are committed as a rational agent to thinking of your freedom and well-being as ‘necessary goods’. So, he argues, you are committed to supposing others must not interfere with your obtaining and possessing freedom and well-being. And this, Gewirth claims, is to suppose you have a right to freedom and well-being simply because you are a rational agent. But, he concludes, that supposition commits you to allowing that all other rational agents have the same right and, therefore, that you must not interfere with their obtaining and possessing freedom and well being.” (Harman, Explaining Value, 62).

Harman continues on to say of this argument that there is “no way to interpret it that avoids equivocation” (62). Harman argues that the equivocation occurs either at the notion of non-interference or of a right, depending on how the argument is interpreted. In either case, the argument founders on the problem of moving from having an individual understanding of non-interference or of a right to that applying to others apart from the individual. Given individual claims about rights on the part of various individuals, “it still does not follow that you have any reason to accept their claim in the sense of being motivated not to interfere with them,” (63).
would likely also be motivational problems even if the contented criminal accepted Korsgaard’s view as being theoretically correct. The problem is what Korsgaard would call a problem of justifying morality. What can we ourselves who accept Korsgaard’s view say that the contented criminal is getting wrong? What “justifying reasons” can we imagine from the point of view of the contented criminal herself that would justify in her own sight the choice to give up a life she likes at great cost to herself? Why should she continue with reflection, and what does she do wrong if she doesn’t?

One is taught always to answer rhetorical questions like these, but the answers here won’t be easy to come by. If the contented criminal refuses to continue reflection to the point where she comes to see that she must give up that practical identity, it is very difficult to see how she could be getting anything wrong with respect to the world as explained by the Modern Scientific World View. There is nothing there to address her as a practical agent (although Chapter 5 will consider if John Post might show us how one might move from nature to normativity after all). Nor is she failing to do right by the claims that other people might make on her. The first-person standpoint is not the second-person standpoint, and nothing in Korsgaard’s view requires that the second-person standpoint of the other be taken as valuable in its own right apart from first-personal considerations. Korsgaard’s view is explicitly first-personal, and the second-person address of the other person must first pass through the first-personal standpoint before it can be cloaked in something that makes it a reason for the person in the first-personal standpoint. “Normative reasons,” one will recall,

---

135 For some forceful reflections on these sorts of problems for a view like Korsgaard’s, see Hare, *The Moral Gap*. 
“always come from reflective endorsement” for Korsgaard. Finally, it turns out, all that she would be getting “wrong” is an ideal of humanity as consistency, as Korsgaard conceives of that. This honestly might not seem like that much to be getting “wrong.” Indeed, it is difficult to say just what is so right about it that it offers a justifying reason for doing something very costly—which, as Korsgaard recognizes—is exactly what someone who follows the obligations of morality must often do.

Now Korsgaard would immediately and vigorously protest that “humanity” is no mere Kantian ideal, dreamed up in a philosopher’s armchair, but just what is crucial about human life as it is lived daily. It’s the very element of choice that underwrites all that we do that is under attack here, and it cannot simply be pushed to the side as if it were no big deal. But there are at least two replies here. First, as just shown above, there is not necessarily any inconsistency in refusing to take the claims some human beings try to address to you into your first-person standpoint as giving you reasons. First personally, you are not them and your access to their consciousness and decision-making process profoundly differs from your access to your own; performatively, this makes all the difference in the world. And, second, even if one were giving up an ideal of consistency (as is not clear, since the reasons on offer from without need not be viewed as relevantly similar to these deliberations I undertake here and now), one is not giving up the ability to make a choice, nor is one making a choice that is incoherent. One might choose a view of humanity where a rigid consistency was not always the highest value, but could sometimes be trumped by other values. There might be a kind of voluntarism where the will might sovereignly choose to forego consistency for a time in order to pursue some other good that is now being set

above consistency. Perhaps one is pursuing a symphony greater than anything Beethoven
dreamt or a cure for cancer or the theory that will unify particle physics or some
Nietzschean ideal that will give the poets something to sing about ages hence. Why should
consistency trump all of that? And, as Robert Nozick has put the point in connection with
what he calls “the immoral man,” “suppose that we show that some X he holds or accepts
or does commits him to behaving morally. He now must give up at least one of the
following: (a) behaving immorally, (b) maintaining X, (c) being consistent about this matter
in this respect. The immoral man tells us, “To tell the truth, if I had to make the choice, I
would give up being consistent.”

At worst, one gives up an ideal of oneself as rigidly
consistent, but perhaps finds a consistent and workable alternative ideal of humanity—at
least for oneself.

Furthermore (and with a nod in the direction of Korsgaard’s transition problem, to
be explored below), even where there are not such lofty ideals involved, but merely the petty
selfishness of someone who just wants more for herself, the contented criminal might
wonder about this “reflection” Korsgaard is urging her to do. Say that the contented
criminal has read her Michael Ruse and her E.O. Wilson and has gotten a little suspicious of
the continuing usefulness of some elements of her evolutionary makeup. Taking into
account the explanations she believes are true, she feels that the shape her reflections take
are inevitably tilted toward “morality,” and would like to correct for this in her reflections.
If it is indeed the case, as Gibbard does not tire of pointing out in an attempt to expose the
efforts of the sociobiologists, that “our ‘design’ may causally explain what we end up doing,

---

but what to do is a different question,"¹³⁸ then the contented criminal, being fairly shrewd, will recognize that this is a double-edged sword. If her biology cannot give her reasons to dismiss morality or to cheat on her husband, neither can it give her reasons not to—even if the “moral impulse” of reflection sometimes feels quite strong. So perhaps the contented criminal will correct for this influence, in the same way that a “moral person” might correct for the impulse to cheat on his wife or abuse his stepchildren, as some evolutionary psychologists suggest we may be predisposed to do.¹³⁹ The contented criminal might consistently view the very demand for consistency as needing to be corrected for in her first-person deliberations so as not to be taken in by forces of evolution that have arbitrarily predisposed one toward an irrational altruism that is not to one’s own benefit today. Perhaps she might view it as a sort of evolutionary “survival” (as the anthropologists would say) that has outlived its use—at least for her.

Lastly, the contented criminal—having read her Nietzsche as well—might also question why this setting of ends and conferring of values should be thought of in a Kantian manner. Along with this, she might think that an Apollonian conception of humanity is just one conception—and that it is competing against other conceptions of humanity and the practical identities that might fit better there. Why should she let it trump her Mafiosa identity that she likes rather well?¹⁴⁰ Or why should she let it trump being a dictator, if she

¹³⁸ Allan Gibbard, “Morality as Consistency in Living,” 161.

¹³⁹ Allan Gibbard, “Morality as Consistency in Living,” 161, cites an utterly depressing study (Martin Daly and Margot Wilson, The Truth about Cinderella: A Darwinian View of Parental Love (London: Weidenfeld and Nicolson, 1998) that argues that step-parents may be genetically predisposed to abuse their step-children.

¹⁴⁰ Seiriol Morgan has argued convincingly, against the “synthetic realists” (or “Cornell realists,” as they are more commonly known), that a Nietzschean “noble realism” would have as much “explanatory power” (the engine driving the moral realism of synthetic realism) as “moral realism.” In other words, the same rebellion that the Kantian higher man explains in terms of the injustice of the upper class, the Nietzschean
has reason to believe that is a viable option? What’s so valuable about this Apollonian self, anyway? It’s not as if it is getting something right. Korsgaard’s Apollonian humanity might be viewed as just one more possible practical identity. Or perhaps—convinced that Korsgaard is right about humanity as requiring consistency—she may want to “lose herself” (people often speak this way) in some Dionysian revelry for a while, getting out from under the thumb of determining judgment. It could make sense to let the Dionysian trump the Apollonian for a while. (Even in this sentence one sees the presumed supremacy of the Apollonian: “It could make sense to let the Dionysian trump the Apollonian for a while.”) But say she became convinced that the Dionysian is the truly human, and that Apollo is Dionysius’ step-n-fetch. “Humanity”—especially in its capacity to set ends and confer value—is no trump card Korsgaard can play here. It itself comes under dispute, and one might find oneself under the sway of another view of “humanity” than the Kantian view Korsgaard likes. Korsgaard’s appeal to Apollo does not carry the justificatory weight her moral philosophy needs. The Apollonian humanity Korsgaard assumes is one conception of humanity, and one the contented criminal might well challenge—especially if she comes to think that it fits rather poorly with the world as she understands it to be explained.

Korsgaard, then—for all these reasons—has not shown how she avoids moral reasons arbitrariness if identity precedes reasons. Nothing she finds—either in her materialist metaphysic or in the demands inherent in taking up the first-personal standpoint where one decides what to do—successfully constrains the endorsement of one identity over another. There is an arbitrariness to the identities one finds oneself with owing to the aristocrat (leaving aside matters of Nietzsche exegesis) might explain in terms of an enervated aristocracy that was insufficiently ruthless. More than one explanatory category can lay equal claim to the “realism” in the manner in which the synthetic realists establish a putative “moral realism.” See Morgan, “Naturalism and Normativity.”
nature of the forces that hold endorsement in its grip: one endorses this identity because of being born in this culture as opposed to that one or because of this evolutionary happenstance or that one. And various “reasons” emerging from these identities move in different directions and support (normatively, according to Korsgaard) different courses of action. But these “reasons” that support different courses of action enjoy the same grounds; none of them is uniquely justified. And the “human identity” Korsgaard appeals to in her attempts to solve the problem of finding reasons that uniquely justify one course of action over another fails to solve the problem. As a result, reasons—which flow out of practical identities in what seems to be Korsgaard’s considered view in The Sources of Normativity—are mired in moral reasons arbitrariness.

C. Setting Ends: Reasons before Identity

As suggested above, Korsgaard’s considered view in The Sources of Normativity may be undergoing a change in her most recent work. The idea of the maxim is increasingly prominent in her work, and the notion of the maxim as a fundamental structure of human thought—a logos, she calls it—sometimes appears to precede all practical identities, including that of being human. In short, Korsgaard may be starting to incline toward prioritizing reasons over identity in a very fundamental way, in what is here being called the form-priority maxims view. This subsection of the dissertation will argue briefly that if Korsgaard is indeed moving in this direction (something she seems to be doing in her latest work), she will meet with no more success than she did in prioritizing identity over reasons, with the result that here, too, Korsgaard’s moral philosophy fails to avoid moral reasons arbitrariness. In other words, the present subsection does not make a claim to authoritative exegesis of an
incipient movement in Korsgaard’s most recent thought, but rather is an attempt to
foreclose a possible direction a moral philosopher inspired by Korsgaard might be tempted
to explore.

Perhaps the place where this incipient movement in Korsgaard’s thought can be
seen most clearly is in a recent interview she gave on the topic of her moral philosophy.
Asked about a criticism often directed toward Kantian moral theories, namely, that they are
tightly confined to the first-person point of view and that they miss the intersubjectivity of
morality, Korsgaard responds (in a passage already quoted above) by defending the public
nature of reasons on her view.

The idea that reasons are private, that they might belong to one person in particular,
seems to assume that we can identify one person in particular in advance or prior to the person’s
reasons themselves. I also hold a kind of constructive view of personal identity, and I
think that what gives us a personal identity is the reasons that we autonomously
adopt for ourselves. And that means that you cannot take for granted the notion of
somebody’s being me in advance of which reasons I have and adopt. My reasons are part of my
practical identity, and I construct my practical identity, and so I do not exist, so to speak, ahead of
my reasons.141

There are a number of tensions in this passage regarding the relationship between identity
and reasons, and it sometimes appears that the endorsement of an autonomous self still
precedes reasons: we “autonomously adopt” our reasons, Korsgaard says. And yet this is
what leads to our various practical identities, and all the problems I have just argued vitiate
the identity-priority maxims. Now perhaps Korsgaard means nothing more than that our
various local practical identities (husband, father, son, volleyball player, musician, and so on)
all owe finally to an overarching human identity that precedes them all. In this case, identity
still precedes reasons and the criticisms of the previous subsection apply.

141 Korsgaard, “Internalism and the Sources of Normativity,” 55; emphasis added, with the
underlining pointing out what I take to be perhaps the most important phrase for the form-priority reading of
Korsgaard on maxims (see also the passages at notes 98-101).
However, at times it sounds as though she might be tempted by a stronger claim—especially when we recall what she said about maxims in the passages quoted above. Korsgaard seems to be claiming that there is truly no identifiable person “in advance of or prior to the person’s reasons themselves.” The person does not exist ahead of her reasons. Reasons would then seem to have some sort of existence prior to or independent of persons. As seen above this would reside in the formal structure of the maxim (the *logos*) that *just is* the structure of thought that makes a human *human*. As a particular animal comes to do an action *for* an end—as it starts to have maxims it follows—there is a necessary structure to this emerging ability to *think*, to *decide* what to do (rather than just doing one thing or another). *Its humanity emerges because it begins to think in the only pattern of thought that makes one human: the maxim.* It has reasons for the first time, and these reasons—or, rather, the maxim structure of thought—constitute the animal as human. This structure of thought that first makes a particular primate *human* of necessity imposes constraints on any being that is human, constraints of following a certain formal structure in which one “considers promoting a certain end by means of a certain act done in a certain way at a certain time and place.”\(^\text{142}\) So it could be that Korsgaard is moving toward saying that while a particular kind of animal exists and does things prior to having an identity, when it makes its first autonomous choice—that is, its first actual choice as a human being—that choice has to have one particular formal structure and cannot but be governed by that structure. As has been seen above, Korsgaard would see this structure as expressing the categorical imperative.

Before developing the central problem regarding Korsgaard’s possible appeal to the maxim-structure of human thought, it will be good to recall a few difficulties argued for above that also apply here. The first thing to note is that here again Korsgaard faces the same problem Gibbard pushes regarding the inadequacy of mere consistency to direct human moral thought. No matter how consistent the system of beliefs, it need not be “moral” unless external inputs into the system constrain it in ways appropriate to what is “moral.” To quote Gibbard again, “If our proclivities go wrong, then we go wrong in our life policies, be they every so consistent.” At most, the maxim-structure of thought only affords consistency as a ground for morality, and (as others have argued and as I have also argued above) that doesn’t seem to be enough to do the work that needs to be done.

Second, making the maxim-structure of thought prior to identity would not answer the contented criminal’s question of about whether or not this maxim structure is merely a shill for other forces. From within the first-person standpoint—even if it is called out and founded in a maxim-structure of thought—it is still possible to ask if one should resist a moral impulse arising from this as being an outmoded evolutionary imposition. Just as the fact (if it is one) that knuckles evolved for (see Chapter 5 for an examination of what might be meant by this) hitting others when vying for a mate doesn’t by itself give one a reason to hit someone, the fact a particular primate evolved that accepted the maxim-structure because such behavior accrued survivability benefits (due to increased in-group cooperation or whatever) doesn’t give one a reason to be moral. Once the contented criminal sees this, moral reasons are seen through. They lose much of their putative authority over her thinking as they are revealed to owe their hold over her to forces that have no claim to

143 Gibbard, “Morality as Consistency in Living,” 146.
moral authority. The contented criminal may be thankful that the maxim-structure of thought which opened her out into the first-person standpoint allows her to distance herself from being dominated by an evolutionary history that makes the categorical imperative attractive, or she may simply find it a matter of little difference if she finds reasons of her own to elevate certain impulses over moral reasons, since whatever “choice” she ends up “making” she might take to be a response to the non-moral forces at play.\footnote{As will be seen in the next section, Korsgaard insists that determinism makes no difference whatsoever to the decisions one makes in the first-person standpoint. This mistaken view will come in for some direct criticism in that section.} And, in any case, neither outcome gets anything right or wrong. If there is a formal lawfulness of the maxim structure of thought but the contented criminal in the first-person standpoint thereby opened up sometimes simply chooses to do something out of step with this lawfulness, what—really—has she violated? As was argued above, it’s difficult to see a viable candidate. The contented criminal who accepts the Modern Scientific World View might well take herself to “see through” any candidate that is advanced as something that does not have authority over her, but is something whose attraction for her or others she may explain in non-moral terms.

But all that is by way of review of problems for the identity-priority maxims view that are also applicable if Korsgaard were to head in the direction of a form-priority maxims view. More fundamentally, however, the form-priority maxims view faces its own unique difficulties with the identity/reasons problem. As was argued above, when Korsgaard made identities fundamental, she ended up having identities that were arbitrary and made moral reasons arbitrary in turn. Identities were endorsed without reasons. Identities need settled, non-arbitrary reasons if they are not be arbitrary. But if Korsgaard seeks to avoid this
problem by making reasons fundamental in the form of maxims, her moral philosophy is again caught in the identity/reasons circle. Maxims need a settled, non-arbitrary identity in the person endorsing them if they are not to be arbitrary. Most of the remainder of this subsection will be devoted to clarifying the problem here, but briefly, the problem facing a Korsgaard-inspired philosopher who developed the form-priority maxims view is that the maxim-structure itself—an act for the sake of an end—would appear to be consistent with all sorts of actions and not only those that are morally right. In order to identify good maxims, there needs to be a rational agent who brings a demand for exceptionless consistency to the maxim. The maxim-structure cannot underwrite reasons apart from a rational agent existing prior to the maxims.145

In The Sources of Normativity Korsgaard appears to see the problem, and it is part of what drives her away from the form-priority maxims view toward the identity-priority maxims view. In the paragraph after a remarkable passage where Korsgaard states that a “good maxim is an intrinsically normative entity,” she states that the possibility of consistently willing a maxim forms a test by which we may determine whether or not something is a reason. This test, she says, “is a test of endorsement.”146 Here the maxim does not so much constitute morality by means of consistency as form the basis for a test of the will of a person, such that if that maxim can be endorsed consistently, then there is a moral reason. So, at the center of things here is a rational human agent who endorses a maxim. The maxim is not

---

145 In Self-Constitution, Korsgaard’s most recent work, she offers a solution to this problem based on the notion of the agent as essentially an activity rather than a state (something that again sounds very Heideggerian). She builds considerable machinery in working up to this solution, drawing on Kant, Aristotle, and Plato. Again, Korsgaard’s book was published as final revisions of this dissertation were underway, and full consideration of her reply would take another 30-40 pages. For now I can do no more than offer my judgment that the critiques I offer here proleptically address her reply. Indeed, not only what I say here, but also what I say about “heautonomy” below strikes me as very apropos.

146 Korsgaard, The Sources of Normativity, 108.
somehow independent from the will of a person with an identity. There is a question of who it is who is doing the willing of this particular act for this particular end, and what that person sees in the end and in the act that makes them worth pursuing. Indeed, just a few pages later Korsgaard confesses that the maxim cannot function on its own as a moral reason: “if it can be willed as a law it is a reason.”

It is important to understand why Korsgaard insists in *The Sources of Normativity* that maxims are not enough: “there is still a deep element of relativism in the system. For whether a maxim can serve as a law still depends upon the way that we think of our identities.” A maxim is merely an act for the sake of a certain end, and both the act and the end require the person’s endorsement if they are to be at all. Indeed, where else could they exist other than persons? After all, Platonism will hardly be attractive to materialists, and Moore receives a stiff challenge from Mackie. Moreover, there are all sorts of impulses vying to be the end for our actions. But none of these is an end until it is endorsed by someone. According to Korsgaard, we humans set ends; none are given to us. But human beings do lots of different acts for lots of different reasons, as Korsgaard recognizes (and this is key): “as I’ve said already, different laws hold for wantons, egoists, lovers, and Citizens of the Kingdom of Ends.” As people’s practical identities diverge so do their maxims—and each of them has the appropriate structure of the maxim: a pairing of act with end. But without a human identity that is settled and unavoidable, the maxim structure by itself cannot

---

147 Korsgaard, *The Sources of Normativity*, 113.


constrain human willing. Thus Korsgaard states that the categorical imperative “simply arises from the nature of the will. It describes what a free will must do in order to be what it is. It must choose a maxim it can regard as law.” Here we see that the will preexists the categorical imperative and the maxim, and its nature forms the test that must be brought to a maxim to determine if a maxim can be regarded as a law. The maxim and its form are not enough. A rational identity must first arise within some animal that will consistently be reflective in considering which maxim to endorse. The identity supplies the consistency through its own capacity to be reflective and its consistent bent to do so. Without the identity of the human being securely in place, any number of maxims can be endorsed consistently. If there is no stable human identity to constrain endorsement, the maxims end up beholden to acts of endorsement unconstrained by a reflective identity of the person who is doing the endorsing. The appeal to maxims thus circles back around to the need for a human identity, and so Korsgaard spends a great deal of effort in *The Sources of Normativity* trying to show that a certain kind of animal does in fact find itself constrained to take up a human identity, understood in her Kantian manner. But it was argued above (section 2, subsection B) that this attempt—putting identity before reasons—succumbs to moral reasons arbitrariness. Either way, Korsgaard’s moral philosophy ends up mired in moral reasons arbitrariness.

But what if one granted Korsgaard the central idea of the form-priority maxims view and allowed that the maxim-structure precedes the human identity and calls it out by virtue of its form? Even granted this, the form-priority maxim view would not be able to do the work Korsgaard needs it to do. The problem here is that what this structure gives rise to is able to transcend the limits of the form that (somehow, as is being granted for the moment)

---

151 Korsgaard, *The Sources of Normativity*, 98; emphasis added.
gives rise to it. The subject ends up being able to set ends and confer value, these not being
given to the subject apart from itself. And, as Allen Wood noted in a passage quoted
above—and is made clear by Korsgaard’s response to the Mafioso problem—the ability
here is not merely that of conferring value in the sense of recognizing that certain
requirements have been met. There is a form of sovereignty here. Humans set value and
it’s not clear what the world of the Modern Scientific World View might offer that would be
capable either of resisting or demanding our bestowal of value (although Chapter 5 will
consider a candidate put forward by Millikan and Post). According to Korsgaard, no moral
value exists apart from humans bringing it into being, and the form of the maxim, but it
seems that once it has called forth a particular animal into a place of being able to recognize
ends as such and endorse them as it chooses, is like a ladder that can be kicked away. The
animal is no longer constrained to choose only maxims that conform to that initial form, but
can now freely choose to follow other maxims—maxims that do not conform to the
lawfulness Korsgaard sees as essential.\footnote{Here, again, it seems that Heidegger may have seen the problems more clearly than Korsgaard. Upon first coming into “the clearing” of thought, where the oppressive closeness of the seamless skein of things that dominates all other animals “falls away,” the very distance from all givenness that gives rise to the Being of things (that is, their being what they are in human reflective consciousness) leaves human beings without guidance. Indeed, in Sartre the very attempt to recognize such guidance can only be a “bad faith” attempt to get oneself off the hook of being abandoned to absolute responsibility (which, interestingly, turns out to be a complete lack of accountability to anything). All of this may be a central reason why Heidegger never developed a way of thinking about ethics (and why Sartre’s ethical thought is so patently inadequate).}

152 Even to show (as would be difficult indeed) that
the maxim-structure initially constitutes the subject as a free, reflective subject is far short of
what Korsgaard would need if she were to pursue the form-priority maxim account of moral
reasons. She would have to show that a particular “lawful” form continued to be necessary
for the particular willing of any agent that had the capacity to appreciate, weight, and act on
the reasons relevant in any particular situation. To say that it is quite unclear how such a story would go is a fair-sized understatement.

Finally, it should be noted that it is also quite unclear how any of this could fit with Korsgaard’s materialist commitments. How could the maxim-form of thought exist somehow independently of any particular person who could think it so as to call out all such persons? Korsgaard would not be likely to countenance some sort of Platonic form—the maxim—that somehow existed of itself. What, then? *Ex hypothesi*, there can be no persons yet in existence to play the needed role. Moreover, since a maxim is an act-end pairing, having the maxim-structure of thought precede and call out persons as agents means that ends would have to exist prior to persons somehow in the material world. But a central tenet of materialism is that there are no final causes in the natural world. Only human beings (or other contingent rational beings) bring final causes into the world. Given the Modern Scientific World View, the maxim cannot exist somehow independently of contingent personal beings such that they can recognize it and be drawn to it as it stands somehow apart from them. In J.L. Mackie’s terms, that would be “queer” if anything would. In short, the Korsgaard has no easy road forward here. Perhaps the best thing would to be to look to mathematics for a partner in the guilt and to say that—somehow—it just does in fact work. This is less than wholly satisfactory.

Korsgaard thus faces serious difficulties if she follows either the form-priority maxims account or the identity-priority-account of the provenance of moral reasons. Not only is the form-priority maxim view is of dubious consistency with materialism, it has a

---

153 “Like mathematics, the maxim seems to stand on its own as something *sui generis,*” Korsgaard might say. “That we don’t understand what this independence looks like doesn’t lead us to question mathematics. Nor should it lead us to question the maxim.”
more pressing problem: No less than the identity-priority maxim view, the form-priority maxim view succumbs to moral reasons arbitrariness, since the mere form of maxims cannot uniquely justify a particular course of action, but leaves us with “moral reasons” that move in divergent directions. The only way to rectify the problems here would be to appeal to the rational agent who preexists the maxims and demands exceptionless consistency from them. But this is a return to the identity-priority maxims view—which I have already argued is flawed. Korsgaard remains caught in the identity/reasons problem—and whichever portion of the circle she emphasizes leads to moral reasons arbitrariness.

In this section of Chapter 4 it has been argued that the identity/reasons problem shows that whether Korsgaard starts with reasons or identities, she lacks the resources to constrain human endorsement, with the result that her moral philosophy fails to avoid moral reasons arbitrariness. The next section will try to expose what might possibly be the deeper context for Korsgaard’s failure to give a satisfactory account of moral reasons.

III. Korsgaard's Kantian Transition Problem

As already noted above, Allan Gibbard finds himself very much in agreement with Korsgaard regarding the starting points of her system, starting points that lead him to call her an expressivist. But Gibbard criticizes Korsgaard for developing a very strong notion

---

154 The original plan for this section of Chapter 4 involved covering two Kantian problems that Korsgaard faces. The first, covered here, involves Korsgaard’s transition problem. This problem is so involved and intricate that issues of time and space started to make it obvious that covering this problem would be enough. There is, however, a second Kantian problem Korsgaard faces regarding the public nature of reasons that could helpfully be viewed in light of Kant’s notion of the sensus communis in the third Critique. In hopes of not making this one chapter into a dissertation by itself, nothing more will be said of this other difficulty.

155 Whether or not Gibbard is right to do so is not of concern here. It seems that a possible take on the matter is that he is right to notice deep affinities—and that these affinities are more basic than the issues
of reflection that is implausible given their shared materialism. Korsgaard starts in the right place, according to Gibbard, properly emphasizing “expressivist” starting points that ground human morality in human activity, but then she takes a wrong turn: “Korsgaard thinks that what follows is an elaborate Kantian package; I think that what follows is something more modest.” On Gibbard’s view, we need to think much more carefully than Korsgaard does about what natural selection could plausibly have given rise to with respect to our moral capacities, and that when we do so we do not find an intricate Kantian philosophy, but something much more rough and ready. Fortunately, Gibbard thinks, what we find there is enough, for the pressures coming to bear on us in natural selection have ensured that we will be wired for moral behavior. And if they did not, then no amount of consistency in our willing would fix the problem. As Gibbard puts it, “our judgments of how to live depend not just on a logic of what to do, but on our proclivities to ‘see’ some eventualities as goals to pursue, and others as dangers to shun. If our proclivities go wrong, then we go wrong in our life policies, be they ever so consistent. Illogic isn’t the only way to be wrong.” Nor will logic be able to set things right if these proclivities lead us astray.

One way of summing up the difference between Gibbard and Korsgaard on these matters is to say that Gibbard brings the third person realm of explanation much closer to that separate them. But Korsgaard intends to be a cognitivist, and Gibbard finally is not (although he wants to sidle up as close as possible). Perhaps one could say that they are different kinds of constructivists, or perhaps one could say that they are different sorts of projectivists, with Gibbard thinking that Darwinism lays out the machinery that does the projecting of plans, while Korsgaard thinking that Kantian philosophy lays out the machinery that does the projecting of practical reasons.

156 Gibbard, “Morality as Consistency in Living: Korsgaard’s Kantian Lectures,” 141.

157 Gibbard, “Morality as Consistency in Living: Korsgaard’s Kantian Lectures,” 146. As argued in Chapter 3, Gibbard’s views have their own problems. It’s not clear that the “proclivities” we might be thought to have in the Modern Scientific World View will be able to carry the moral weight Gibbard needs them to carry.
the realm of our lived choices than Korsgaard is willing to allow, with Gibbard grounding our moral life in *a posteriori* matters of how things have actually played out with respect to the relevant evolutionary forces, such that the substantive “proclivities” we have play a critical role. Indeed, Chapter 5 will consider the view of John Post, which makes *a posteriori*, evolutionary considerations even more central than they are for Gibbard. But for Korsgaard, the “proclivities” merely provide “incentives” that the autonomous human makes into reasons through reflective endorsement or through the actual formation of maxims that already express such autonomous endorsement in their very form. So when Korsgaard worries about a “specter of unmasking” that threatens morality when we regress on the justifying conditions for any particular choice to “do what is right,” she turns to a procedure to stop the regress—a form of practical reasoning that is correct in virtue of its form. Gibbard, on the other hand, argues that the “why-stopper” must come in the form of substantive principles, principles arising from our evolutionary history. As Gibbard points out, Korsgaard habitually accuses any such appeal to substantive principles as being part and parcel of an objectionable metaphysical moral realism, a charge Gibbard is quick to try to deflect: This issue of substantive “why-stoppers” “is a distinct issue from the one that pits Korsgaard’s expressivism against a gratuitous, metaphysical moral realism.” In short, Gibbard and Korsgaard each want to hold that their view better comports with what Gibbard called the “Galilean core” of the Modern Scientific World View, each thinking that

---

158 Though it has already been seen that Korsgaard fails to avoid the need to go back to these “sustantivalist” starting points.

159 The phrase “specter of unmasking” is from Gibbard, “Morality as Consistency in Living: Korsgaard’s Kantian Lectures,” 140.

160 Gibbard, “Morality as Consistency in Living: Korsgaard’s Kantian Lectures,” 142.
the other is betraying it in some way. Gibbard sees that Korsgaard’s elaborate Kantian machinery is obviously metaphysically suspect from within the Galilean metaphysic, while Korsgaard sees that Gibbard’s substantive moral proclivities—if they do not fail to halt the regress altogether—are freighted with metaphysical baggage Gibbard needs to smuggle into his account to get it to address moral problems at all.

Gibbard’s difficulties have already been addressed in Chapter 3, and it seems that Korsgaard’s critique is on target in many ways. At the same time, Gibbard has also identified a real point of tension in Korsgaard’s moral philosophy, and that will be the focus here. In short, Gibbard is on to something very much like the “transition problem” that worried Kant mightily in his *Critique of Judgment*. And he was right to be concerned. The transition problem is a severe difficulty for Kant’s philosophy, and Kantians after him—including Korsgaard—have continued to struggle with it.\(^{161}\) In Korsgaard’s philosophy, the transition problem is seen most clearly in her doctrine of the two standpoints, especially regarding questions concerning how the first person standpoint of decision and justification is related to the third person standpoint of explanation and prediction. Before getting into these issues in Korsgaard’s philosophy, however, it will be good to sketch the outlines of the transition problem with brief reference to Kant.\(^{162}\)

---

\(^{161}\) Although it will not be argued here, upon continued reflection on what Kant called the “transition” problem, it seems that the problem is not really just a Kantian problem, but a problem for anyone who gives the cognitive weight of explanation wholly to materialism, and yet does not wish to eliminate (by theoretical fiat, for certainly there is an *apraxia* problem here) lived human experience. Thomas Nagel is a contemporary philosopher who acknowledges the wider problem more than most philosophers. The Kantian solution to that wider problem is what Kant called “heautonomy” (which will be sketched presently).

\(^{162}\) The work that follows draws on previous work done in Bradley N. Seeman, “What If the Elephant Speaks? Kant’s *Critique of Judgment* and an Übergang Problem in John Hick’s Philosophy of Religious Pluralism,” *International Journal for Philosophy of Religion* 54, no. 3 (December 2003), 157-74; and, prior to that, Bradley N. Seeman, “Hermeneutics, Heautonomy, and Hope in Kant’s *Critique of Judgment*,” unpublished mss. (December 2001).
By the time of the third Critique, Kant realized there was a severe difficulty facing his critical philosophy, namely that of bridging or transitioning between his philosophy of nature and cognition in the Critique of Pure Reason and his philosophy of freedom and practical desire in the Critique of Practical Reason. As Kant put it in the published introduction to the Critique of Judgment,

Now although there is an incalculable gulf fixed between the domain of the concept of nature, as the sensible, and the domain of the concept of freedom, as the supersensible, so that from the former to the latter (thus by means of the theoretical use of reason) no transition is possible, just as if there were so many different worlds, the first of which can have no influence on the second; yet the latter should have an influence on the former, namely the concept of freedom should make the end that is imposed by its laws real in the sensible world; and nature must consequently also be able to be conceived in such a way that the lawfulness of its form is at least in agreement with the possibility of the ends that are to be realized in it in accordance with the laws of freedom.—Thus there must still be a ground of the unity of the supersensible that grounds nature with that which the concept of freedom contains practically, the concept of which, even if it does not suffice for cognition of it either theoretically or practically, and thus has no proper domain of its own, nevertheless makes possible the transition from the manner of thinking in accordance with the principles of the one to that in accordance with the principles of the other.¹⁶³

Kant has identified a potential flaw in his system of critical philosophy: Given that (1) the domain of practical philosophy’s faculty of desire is the supersensible, “known” only insofar as we postulate our internal principles of action that are needed by the practical reality of our moral action and only possible in complete isolation from vitiating motives of the sensible realm, and given that (2) the sensible domain of the faculty of cognition is constituted according to a seamless lawfulness imposed by the categories of understanding and thus

cognitively closed to any irruption by reason’s postulate of freedom, how can Kant bridge the gulf between these two realms so that reason can freely make its end “real in the sensible world”? One commentator states that it can almost appear that there is an antinomy between the first and second Critiques that Kant sets out to reconcile in the third Critique.\(^{164}\)

Kant’s answer to the bridging (or transition) problem hinges on what he called the heautonomy of the faculty of judging, with reflecting judging legislating not a domain so much as constituting itself according to its own rules in an effort to reach for objectivity.\(^{165}\)

Etymologically, Kant means to emphasize the self-reflexive character of judgment’s legislation by attaching the Greek definite article “\(hē\)” to the pronoun “\(auto\).” The resulting

\[^{164}\] A.C. Genova, “Kant’s Complex Problem of Reflective Judgment,” *Review of Metaphysics* 23, no. 3 (March 1970): 475; and A.C. Genova, “The Purposive Unity of Kant’s Critical Idealism,” *Idealistic Studies* 5 (1975): 181. Genova goes on to deny that there is a real antinomy, because (1) it is not a question of whether or not freedom exists, but how it harmonizes with nature, and (2) even that lack of harmony would not impact the validity of their legislations in their respective domains. Genova goes on to state that “It is not really an opposition or antinomy between theory and practice, but a heterogeneity or gap that Kant would like to bridge (but not eliminate), a discontinuity in our mode of thinking about the principles of nature and freedom” (Genova, “Kant’s Complex Problem,” 476). Genova here accepts Kant’s heautonomous deontologization of the problem of reflecting judging, a move I challenged in “Hermeneutics, Heautonomy, and Hope.”

\[^{165}\] I wish to reemphasize that I am not claiming to give authoritative Kant exegesis. What will be of most importance here is not the term “heautonomy” (though I think—not without trepidation—that it is the right one for the self-reflexivity central to Korsgaard) nor talk of a Kantian “transition problem” (though I think—with less trepidation—that this is the right way to characterize a problem of relating the practical to the theoretical that Korsgaard does in fact have), but the idea I am trying to unpack in relation to those terms—whether they capture Kant’s own ideas or not. That said, I do find the business of trying to understand Kant both fascinating and valuable.

Talking about heautonomy seems right in relation to Korsgaard because she is trying to identify a kind of “purposiveness” outside of the subject—namely, the purposiveness of other human beings. They are more than just natural objects, but have a kind of purposiveness in being ends for themselves that must be regarded by me as having a purposiveness we share in common if the moral life is to take others into account. It is on this point that I take a crucial aspect of Korsgaard’s thought to turn: “You make yourself an end for others; you make a law to them. But if you are a law to others in so far as you are just human, just someone, then the humanity of others is also a law to you. By making you think these thoughts, I force you to acknowledge the value of my humanity, and I obligate you to act in a way that respects it,” (Korsgaard, *The Sources of Normativity*, 143). There must be a purposiveness—a value I cannot help but recognize—apart from myself. I am made to think about this external purposiveness. Kant’s own way of talking about this in the Critique of Judgment is in terms of heautonomy. The problem that both Kant and Korsgaard face here is that this just pushes the transition problem back to a new level: will you stay within the ambit of the architectonic or not? Both Kant and Korsgaard finally refuse pierce the circle of the architectonic, I believe, creating real difficulties for themselves. That, at any rate, is what I will argue.
“heauto” takes on a necessary self-reflexivity, as in “he patted himself on the back.” Kant’s newly minted word strongly emphasizes that reflecting judging constitutes only itself, having no domain over which it presides. In other words, there is a turning of judgment upon itself so as to create what Korsgaard calls a standpoint, specifically, the standpoint of the agent who decides what to do or who explains the world. The agent, or perhaps better, the form of the judgments of a being who reasons, gives rise to a certain purposiveness with respect to the judgments themselves. The Kantian principle for the faculty of judging is a purposiveness without a purpose, whereby the faculty of judgment “gives a law only to itself, and not to nature.” That is to say, the faculty of judgment operates heautonomously when it legislates over itself with the principle of the purposiveness of nature: “one must call this legislation heautonomy, since the power of judgment does not give the law to nature nor to freedom, but solely to itself.” In this, a kind of subjectively unavoidable objectivity is

166 These etymological insights owe to Juliet Floyd’s admirable paper, “Heautonomy: Kant on Reflective Judgment and Systematicity,” in Kant’s Aesthetics, ed. Herman Parret (Berlin: de Gruyter, 1998), 192-218, here 205. Floyd herself credits these insights to conversations with Nikolas Pappas and Jennifer Roberts.

167 Henry Allison goes so far as to call this principle of the purposiveness of nature the “category” of reflecting judging. See Henry E. Allison, Kant’s Theory of Taste: A Reading of the Critique of Aesthetic Judgment (New York: Cambridge University Press, 2001), 207.

168 Immanuel Kant, “First Introduction,” in Critique of Judgment, 27-8/20:225; italicized emphases added (see also 72/5:185-86). In full, this passage reads as follows:

this autonomy is not, however (like that of the understanding, with regard to the theoretical laws of nature, or of reason, in the practical laws of freedom), valid objectively, i.e., through concepts of things or possible actions, but is merely subjectively valid, for the judgment from feeling, which, if it can make a claim to universal validity, demonstrates its origins grounded in a priori principles. Strictly speaking, one must call this legislation heautonomy, since the power of judgment does not give the law to nature nor to freedom, but solely to itself, and it is not a faculty for producing concepts of objects, but only for comparing present cases to others that have been given to it and thereby indicating the subjective conditions of the possibility of this combination a priori (Immanuel Kant, “First Introduction,” in Critique of Judgment (New York: Cambridge University Press, 2000), 27-8/20:225; italicized emphases added).

That there is a case of heautonomy here and not merely the legislation appropriate to the domain of freedom owes to the fact that there is a domain that is being constituted in this case, namely, actions within the world of
constituted for oneself. As Henry Allison rightly observes, Kant’s claim “is not that nature is purposive, that is, that we have some sort of a priori guarantee that it is ordered in a manner commensurate with our cognitive capacities and needs. Nor it is even that we must believe it to be purposive in this sense (which is basically Hume’s position). The claim is rather that we are rationally constituted to approach nature as if it were so ordered.” Here, as elsewhere, Kant uses ontology for his purposes, but finally places it under an epistemological and heautonomous “as if.” Kant is clear. Purpose cannot be known to exist outside the standpoint of our own act of judging. Korsgaard captures the essence of the idea: “the way we conceptualize the world, the way we organize it into a world of various objects, guarantees that it will appear to be teleologically organized at the level of those objects.” Purposiveness exists only as our judging sets a rule for itself, and does not exist somehow independently of us in the natural world.

But it’s difficult to see how Kant has rendered the transition problem tractable here, for judgment has merely circled back on itself, and Kant makes it quite clear that

---

169 Allison, Kant’s Theory of Taste, 39.

170 Seeman, “Hermeneutics, Heautonomy, and Hope.”

171 Korsgaard, Self-Constitution, 39. Or, again: “Teleological thinking need not be grounded in a claim about the world. It may be grounded in a claim about how human beings conceptualize the world” (38).
purposiveness does not reach beyond the ambit of that circle. The question, then, is about the status of a purposiveness that applies only to the judgment and only through the judging itself (i.e., is self-reflexive), such that it gives no indication of a reality apart from that of its own judging. Poised between nature and freedom, as Kant puts it, such heautonomous judging has no cognitive, explanatory purchase; yet it is not of morality, where there are no objects, but only the spontaneity of will. Instead, it operates as if the claims it made could apply to judgments of objects. It’s simply that the subject is constituted so that it cannot help but judge as if there is a purposiveness independent of the judgment, a purposiveness to which the judgment is adhering. As one commentator notes, Korsgaard’s “as if” here reflects her claim that what we accept for practical purposes carries no theoretical claims with it. To use Korsgaard’s terminology, it has none of the cognitive, explanatory power of the third-person standpoint; indeed, even the heautonomous judgment itself admits of explanation. Even while our capacity of reflective judging cannot help but judge as if there is a purposiveness apart from itself, none of this has cognitive, explanatory force. Moreover, it must be admitted that judging must itself be explicable in event causal terms (cognitively, in determining judging), and that the judgments of purposiveness occur in the

---

172 The following paragraph wades into the murky waters of Kantian terminology—perhaps too deeply for some, and not deeply enough for others. Unfortunately, I don’t see a way of either avoiding the Kantian argot altogether, or of going as deeply into it as it would deserve if this dissertation were an attempt at Kant exegesis. I am trying here only to develop a suggestive connection that might exist between a central part of Korsgaard’s thought and a rather obscure—though I think absolutely crucial—aspect of her Kantian heritage. I believe there is a linkage, but it is not easy to draw it out succinctly. Nothing with Kant works that way.

173 J.B. Schneewind, “Korsgaard and the Unconditional in Morality,” *Ethics* 109, no. 1 (October 1998): 43. Paul Guyer makes much the same point: “As Korsgaard frequently insists… Kant’s argument is not to be understood as an argument in theoretical metaphysics, which uses the concept of rational being or agency to derive the consequences of the alleged fact of our rationality; rather, it is always to be understood as an argument made from a practical point of view, that is, an exploration of the consequences of our choosing to conceive of ourselves as rational beings, though no theoretical explanation can be given for the necessity or even the possibility of such a choice,” (Guyer, “The Value of Reason and the Value of Freedom,” 27.
way they do because of those causal factors. But none of this shows how in a spontaneous
willing a supersensible moral law of autonomy can “manifest itself in the sensible world
without infringing upon its laws.”\(^{174}\) Or, as it might be put in Korsgaard’s terms, Kant has
not shown how justification can find a place in the world of non-normative, event-causal
explanation. Such purposiveness cannot change the course of things in a world seamlessly
explained (at least by promissory notes) in non-normative terms within the Modern
Scientific World View, nor can it carve out space for beings who respond to matters in that
world without being determined by them.

Korsgaard, I want to suggest, faces much the same difficulty as Kant at this point,
for there is no clear sense of how the first person standpoint of practical decision can find
room in a world where cognitively everything may be accounted for in terms of the
materialist, event-causal explanations proper to the business of the third person standpoint.
Thus in Korsgaard we hear echoes of Kant:

as rational beings we may view ourselves from two different standpoints. We may
regard ourselves as objects of theoretical understanding, natural phenomena whose
behavior may be causally explained and predicted like any other. Or we may regard
ourselves as agents, as the thinkers of our thoughts and the originators of our
actions. These two standpoints cannot be completely assimilated to each other, and
the way we view ourselves when we occupy one can appear incongruous with the
way we view ourselves when we occupy the other.\(^ {175}\)

Finally, one’s own moral behavior (and everyone else’s) may be given some materialistically
acceptable, event-causal explanation in the third person standpoint.\(^ {176}\) One’s own behavior

\(^{174}\) Immanuel Kant, *Critique of Practical Reason*, trans. Mary Gregor (1788; New York: Cambridge
University Press, 1997), 38.

\(^{175}\) Korsgaard, *Creating the Kingdom of Ends*, 377-78.

\(^{176}\) When referring to event-causation in this context, I do not mean to be saying anything about the
indeterminacy of our current best subatomic particle physics. The indeterminacy of quantum events is
swallowed down whole—just so long as we’re talking about events rather than agents.
is explained without remainder by causal factors available in non-normative explanatory terms amenable to the Modern Scientific World View, and the person who accepts this also accepts that the behavior of others may likewise be explained in these terms—although, due to various contingencies, that other human animal happens to have been shaped by a different set of events than those which have determined oneself. This itself would already seem to be a prime instance of moral reasons arbitrariness, where one sees that one’s own grounds for acting are no better (or worse) than those of any other agent, since that person and oneself are both determined in their “moral” behavior by non-moral events.

Far from it, however, as Korsgaard sees matters. In fact, so far from this being an instance of moral reasons arbitrariness, Korsgaard holds that it is completely irrelevant to the issue of the justification of morality. Whereas the causal story explains the behavior, the explanation does not address (except in some rather bland, factual ways) the justificatory issue that faces the moral agent at the moment of deciding what to do as an agent. An explanation of a behavior is simply the wrong kind of “reason” to address the moral agent as such. From this first-person standpoint, matters seem quite different from how they are in the third-person standpoint where all is explained cognitively in non-normative, materialist terms. “From a third-person point of view, outside of the deliberative standpoint, it may look as if what happens when someone makes a choice is that the strongest of his conflicting desires wins. But that isn’t the way it is for you when you deliberate. When you deliberate, it is as if there were something over and above all of your desires, something which is you, and which chooses which desire to act on.”

The solution for Korsgaard here, as for any good Kantian when faced with any attempt to push an

177 Korsgaard, Sources of Normativity, 100; italics in original, underlined emphasis added.
antinomy about how things stand “out there” in the world of the things themselves, is to take refuge in the phenomena/noumena distinction. Here it takes the form of two balanced “as ifs”: the “as if” appropriate to the third-person standpoint where everything is seamlessly explained in the non-normative, event-causal terms of materialism; and, on the other side, the “as if” appropriate to the first-person standpoint of the agent who must nonetheless decide what to do as a moral agent faced with a justificatory question of what is the right thing to do. The “as if” qualifying all the materialist, event-causal explanations appropriate to the third-person standpoint, means that finally one cannot take those explanations to give us knowledge of how things are; room is thus opened up for a hope or faith from within the first-person standpoint that is of the essence of our practical lives: “The idea of intelligible causality is a practical conception, and our belief in it is an article of practical faith. It is not supposed to be theoretically employed, and it cannot be used to explain anything.”\(^{178}\) So, in the human practical life, there must be this faith that things are not as they seem from the third-person standpoint of seamless non-normative, materialist explanation. At the same time, the “as if” of the first-person, practical standpoint ensures that the freedom crucial for the agent who is deciding what to do can be disregarded without explanatory loss from within the third-person standpoint of non-normative explanations acceptable within the Modern Scientific World View.

It is interesting that in the same essay in *Creating the Kingdom of Ends* where Korsgaard emphasizes that belief in intelligible causality is “an article of practical faith” integral to the first-person standpoint, she also emphatically downplays the importance of any such belief for morality: “The point is not that you must believe that you are free, but that you must

\(^{178}\) Korsgaard, *Creating the Kingdom of Ends*, 183.
choose as if you were free. It is important to see that this is quite consistent with believing
yourself to be fully determined.” Korsgaard’s point here is that even if a person believes
all of her actions are determined, she is still thrust into a position where she must
nevertheless decide what to do. The fact that she is determined does not make any practical
difference as she is deciding what to do, even though she takes that decision to be the
outcome of forces beyond her control. She will not see which way those forces will carry
her until she decides what to do and responds to whatever reasons she will find persuasive.
“But if we are able to act exactly as we would if we were free, under the influence of the idea
of freedom, then we are free.” Even if she knows that an über-psychologist constantly
predicts what she will do in advance with one hundred percent accuracy, it makes no
difference to her in the standpoint of the agent who must choose what to do all the same.
Thus there seems to be no need for practical faith in the efficacy of our free choice; for
whether we have such faith or not, we cannot help but decide what to do from within a
first-person standpoint. Even if one could know that Skinnerian behaviorism gets things
exactly right with respect the world as it is explained third-personally, it would not affect what
one does when deciding within the first-person standpoint.

There is a tension here in Korsgaard’s thought about the relationship between the
first and third-person standpoints, a tension indicative of a transition problem in her
thought. What connections, if any, are there between the explanations of the third-person
standpoint and the justifications of the first-person standpoint? Can they really be
hermetically sealed off from one another? And what is the relationship between the causes

179 Korsgaard, Creating the Kingdom of Ends, 162; emphasis in original.

180 Korsgaard, Creating the Kingdom of Ends, 176.
operative in explanations and the reasons that an agent can appreciate as offering support for an action and which she can weigh to decide which course of action she should take—which one is justified (as Korsgaard puts it)? Can an agent accept both stories and still have a unified story, or must the agent somehow bifurcate herself so that her acceptance of the one story does not intrude upon her acceptance of the other story? The remainder of this section will advance three difficulties related to this transition problem in Korsgaard’s moral philosophy.  

First, when Korsgaard wavers between holding that a practical faith in free causality is needed for morality, or else holding that this can be laid aside as irrelevant, it is indicative of a dilemma regarding her transition problem. On the one hand, Korsgaard realizes that she cannot allow that practical faith in free intelligible causality is needed for morality, for if it is, then morality is in trouble because this faith flies in the face of Korsgaard’s materialism, with its requirement that everything ultimately arise from event causal forces tied to non-personal stuff and (at least in principle) explicable in those terms. In short, one would need to hope that the materialistic scientific explanation were false, but this barest hope would be in the teeth of cognition and also it would illicitly allow the practical and the 

---

181 There are more than three problems relevant to Korsgaard’s transition problem, but this section contents itself with offering an overview of three of them. Other problems include: (1) how can anything constrain this freedom that operates in complete isolation from all facts, (2) whether any fact that is explanatorily relevant will determine the will and render it ineffectual in determining itself if it has any influence on the will, and (3) the problems related to the incommensurability of the two standpoints (see Korsgaard, Creating the Kingdom of Ends, 204, 206). These problems would bear further thought, but matters of length have dictated putting off consideration of these interesting problems.

182 Of course this is not to say that Korsgaard accepts some sort of Newtonian mechanism; she certainly knows that the current best physics is indeterministic. However, indeterminate events are not, simply in virtue of that indeterminacy, something other than events—even if the nature of the events is opaque. I mean to be placing emphasis here more on the events in event causation than on some sort of deterministic causal closure. The key point, for the materialist, is that all the events—determinate and indeterminate—be non-personal. In all the events there must be nothing “spooky” or “queer” (as some materialists put it).
theoretical to intrude on one another. Moreover, materialists like Korsgaard would be in the unenviable position of having to hope that their own best metaphysical explanations were false in order to enter into the practical standpoint. Making morality vulnerable to the theoretical determinations of third-person explanations in this way ends up pitting morality against cognition—something Korsgaard is (rightly) loath to do—and leaves morality on a very tenuous footing.

On balance, Korsgaard inclines toward grasping the second horn of the dilemma, arguing that the explanations advanced within the third-person standpoint are irrelevant to the agent who must decide as if she were free no matter what explanations are advanced. Korsgaard appeals here to something very much like the Kantian notion of "heautonomy." The very ability to enter into the first-person standpoint involves self-reflexively providing the domain over which practical reasoning legislates in the very exercise of the capacity of judging what to do. The judgment and the domain over which it legislates arise at the same time, and nothing about the course of what is happening with respect to the world as it is explained matters to this heautonomous judgment of what to do. If this is right, then Korsgaard is taking roughly the same route Kant finally took in struggling with the transition problem in the third *Critique*.

---

183 Paul K. Moser (personal communication, 6 June 2008) rightly points out that a kind of fictionalism threatens here, where the self-reflexive judgments have no cognitive value. Though Korsgaard would vigorously object, I work below to establish this sort of problem as a real one for Korsgaard, though I do not use the language of fictionalism.

184 This is at least similar to Kant's pattern of thought in the *Critique of Judgment* (see the passages connected with notes 163 and 168 above).

185 It could be either the practical identity or the structure of the maxim itself that operates in this heautonomous manner in Korsgaard's thought. As noted above, some of Korsgaard's recent writings seem to open up the latter position.
The sharp point of this horn of the dilemma is that deciding what to do within the first-person standpoint cannot be so neatly insulated from what a person accepts as an explanation within the third-person standpoint, for the third-person standpoint inevitably has a first-personal dimension that insinuates the explanations one accepts into the process of deciding what to do within the first person standpoint. In approaching this problem, it is helpful to consider how an explanation looks from the first-person standpoint of someone who accepts the sort of third-person explanations permitted within materialism. Her acceptance of a particular explanatory story brings it over into her decision process in the first-person standpoint. For the same agent who is making a decision from within the first-person standpoint is also the one who believes certain explanations to be true from within the third-person standpoint. After all, it is still the agent herself who explains matters in the third-person standpoint, and at the end of her explanatory efforts she endorses the reasoning process she has followed and believes her explanation in the first-person.

The agent’s endorsement of an explanation commits her—first personally—to any limits that explanation imposes on the ways in which she can understand her own process of deciding what to do within the first-person standpoint. Some things will make sense, and others will not. First, one should note, for one who accepts a materialist, event-causal explanation, free intelligible causality as Kant thought of it doesn’t make sense. Such a person will have to settle for what Kant called that “wretched subterfuge” of compatibilism, or else embrace determinism outright.\textsuperscript{186} The standpoint she adopts will be believed by her

\textsuperscript{186} Immanuel Kant, \textit{Critique of Practical Reason}, 80-81 (5:95-6). Robert Kane has attempted to advance an alternative way forward for materialists. Obviously a full treatment of this lies well beyond the scope of this dissertation. Suffice it to say that Kane’s incompatibilist system seems to be basically a large dose of compatibilism joined to a quantum indeterminacy. It’s difficult to see how this could be an account of free will—especially at the point of its incompatibilism.
to be illusory; perhaps she cannot help but take up the standpoint of having to make a
decision, but the reasoning she undertakes within this standpoint will stand exposed as being
determined by subterranean forces. It is not difficult to see how this might undermine the
ability of those reasons to carry force for her as an agent. She may well think: “Well, this
costly moral behavior seems to be what I have the most reason to do after thinking about
things in light of the categorical imperative; but I know that my finding this reasoning
persuasive owes to certain non-moral contingencies that are meeting at this point of
decision. What I call my ‘reasons’ for this costly behavior manifest the ways in which these
causal forces grip me.” But now I can see that there are reasons for me to wonder whether
I should follow those ‘reasons’, since I see them for what they are—a way in which non-
moral contingencies grip me. Faced with the costs of doing what is ‘moral’, and given the
fact that my aversion to those costs has exactly the same basis as these supposed reasons,
perhaps I should consider what I really do have reason to do.” So the agent might well
reason within the first-person standpoint as she factors in what she believes about how
things are to be explained and what that means for her. A line of reasoning like this would
indeed address her as a first-person agent who must decide what to do. It makes perfect
sense to factor in what one believes explains the reasons one has to do something into the
decision one makes about what to do. And the explanatory story she accepts may well give
her reasons to decide differently than if she were able to accept that she really is a free agent,
or that reasons really do track with some inherently moral reality of which we are a part,

It should also be noted that I don’t mean to be taking any hard stance on this issue myself. Indeed, I
have not myself come to hard and fast conclusions about these matters.

187 Allan Gibbard helpfully draws a distinction between “internalizing” a norm and being “gripped”
whether we like it or not. To ask her simply to disregard the explanatory story she herself accepts is to demand an incoherent bifurcation of herself where one act of acceptance is irrelevant to an act of acceptance made by another “part” of herself.\textsuperscript{188} But since she is a whole person, when she accepts a materialist explanatory story, that story insinuates itself into the first-person standpoint in such a way as to begin to undermine reasons; for she accepts an explanatory story in which the conflicting reasons of other people stand on the same ground as her own, with both owing to an evolutionary history that is unresponsive to moral reasons.\textsuperscript{189}

Something else it will no longer make sense for her to accept upon coming to believe in materialism is that there is anything that everyone should in fact do with respect to the world as it can be explained cognitively.\textsuperscript{190} No one who lived one way as opposed to another could be thought to be getting something wrong with respect to the world as it can be explained. Moral error would have to be shunted into a different category than explanatory error, and this category lacks the same cognitive teeth. Someone who rejects consistency as uniquely appropriate to her first-person decision process does not, strictly speaking, get anything wrong. It’s difficult to see what mistake she would be making with respect to what can be known of the world if she follows Nietzsche or Ruse instead of Kant. Rather she

\textsuperscript{188} Related issues have been explored in Bradley N. Seeman, “Evangelical Historiography beyond the ‘Outward Clash’: A Case Study on the Alternation Approach,” \textit{Christian Scholar’s Review} 33, no. 1 (Fall 2003): 95-124, especially, 110ff.

\textsuperscript{189} Indeed, in accepting this materialist story, she might even come to wonder if evolutionary history is now handing her a golden opportunity where she has seen her way through to the forces that are actually operative in her “moral” behavior and is able to act in ways that may now be more evolutionarily beneficial. Chapter 5 will take up these matters more fully.

\textsuperscript{190} John Post would challenge this. Chapter 5 will attempt to show that there are problems in John Post’s attempts to move from nature to norms.
chooses to weight different factors more heavily within the process of deciding what to do, and she rightly sees that no decision she makes will align her more correctly with respect to how things are in the world. Or, to put matters more carefully, she may well come to believe that immoral behavior may be better for humanity as a whole (perhaps the human race will only survive if the weakest are deprived of the ability to reproduce); or perhaps she will believe that a concern for “humanity as a whole” really has no precedence over a concern for the well being of baby seals or her purchase of a new Lexus, so long as enough cooperation from others can be manufactured so as to serve her interests (that is, so long as it is viable). As a number of commentators point out with regard to Korsgaard’s thought in *The Sources of Normativity*, my seeing that your reasons will give you reason to do something in the same way that my reasons will give me reason to do something, is quite different than believing that your reasons give me reason to do something, and vice versa. The point here is that, having believed the metaphysics of the Modern Scientific World View, there is nothing outside of her that she must believe herself to be violating with respect to any moral behavior. If she ignores the reasons that someone else attempts to address to her, there is nothing in the world as it is thought by her to be explained that she is violating. Indeed, she does not even violate the person whose reasons she ignores, persons being explained in wholly materialist terms.

At this point one encounters a central problem for Korsgaard’s heautonomous attempt to solve her transition problem; for in attempting to solve the problem by an appeal to the self-reflexive, insular nature of the first person standpoint, where the subject not only makes a judgment but projects the domain over which she judges, Korsgaard can at best

---

191 See the passage at note 46 above.
vouchsafe the value of the subject’s own humanity (reflective nature) for the subject herself (that is, subjectively). The subject may not be able to fit herself into her own explanatory story as she actually goes about deciding what to do, but she is never in that situation with anybody but herself. All the other beings around her are explained in the third-person account she accepts, and (given the truth of materialism) she simply does not face the same dilemma with respect to the behavior of others—for she is not in their first-person standpoint. She is not inside their heads experiencing their first-person deliberations as a tension for herself. She experiences only herself (if even that, some materialists would argue) in that manner. The tension that exists for her as she makes a decision in the first-person standpoint, simply does not exist as she considers others; their first-person standpoint is not available to her to create any tension for her. Indeed, from her own vantage, their first-person standpoint—their humanity—may well look illusory. After all, their behavior can (or must, in principle, at least) be explained without remainder from the third-person explanatory standpoint, and the materialist’s explanatory story does not demand that the first-person standpoint of others must be taken seriously within one’s own first-person standpoint. Nothing has bridged the “chasm” Kant saw between the world as known and the world as experienced, and the heautonomous self-reflexivity of the first-person standpoint at best leaves the subject herself in her own standpoint and with all others on the other side of the chasm.

The most that might be said, then, is that such a subject must value her own humanity and that she violates her own supposed humanity in choosing inconsistently with herself. But even this looks highly suspect. For starters, it will be very difficult to get the

---

192 See the passage at note 46 above. The argumentation given here explains why this problem dogs Korsgaard’s philosophy.

193 See note 132 above.
supposed contradiction off the ground when there are no others who must be taken
seriously in their own first-person standpoint. Korsgaard’s strategy of reflective regress to
the “humanity” shared by all practical identities has not succeeded in bringing them over
into the subject’s own first-person deliberations. As Michael Bratman notes, “the iteration
of the demand for reflection by itself does not seem yet to show why the resulting
endorsement cannot be limited to one’s reflectiveness on this particular occasion.”

Moreover, it is not clear what standing the subject’s own “humanity” might have within the
materialist, event-causal explanations offered within the third-person standpoint, any such
elements of the first-person standpoint being explained by evolutionary, biochemical, and
ultimately quantum mechanical terms (or whatever the latest physics will countenance), so
that at best the first-person terms supervene on the third-person explanations—much
the way that the brittleness of glass supervenes on chemical properties. In this way at least
such explanations would have something substantive “underneath” them, such that they
genuinely track something that is really happening. In the same way, the tendency to see
“humanity” in oneself would have arisen from events that could be explained in materialist
terms. And if someone does not see herself in terms of such a reflective identity, she does
not thereby get something wrong. It is simply that something different happens to grip her
behavior than what happens to grip the person with the reflective identity. Someone who
accepts materialist explanations of things will not find anything that is left in terms that may
be recognized from within the first-person standpoint. From within that standpoint, there
is nothing left to address her authoritatively with a reason “of the right kind.” Even her
own “humanity” is stripped away. She may be unable to avoid feeling gripped by it, but this

194 Bratman, “The Sources of Normativity,” 704. See also Raymond Geuss, as quoted at note 46
above.
bit of behavior may be neatly explained away—and those who choose to do so cannot be shown to be getting something wrong and need not think of themselves as having done so.

But perhaps none of this makes any difference. In a remarkable passage, Korsgaard offers a little thought experiment designed to show that even believing (whether correctly or not) that we are determined in our moral behavior makes no difference within the first-person standpoint. Imagine you are hooked up to a little device that works through your own thought processes to determine your behavior. “The device is not going to bypass your thought processes, however, and make you move mechanically, but rather to work through them: it will determine what you think.” Korsgaard clearly means to capture some compatibilist intuitions here. Everything you do will be a result of the device working to shape your own thought patterns, so that when you decide to go to the store it is because your thoughts have led you to do so. It’s just that those thoughts themselves were determined by the device implanted in your brain. Call this the “determinism device.” “The important point here,” Korsgaard asserts, “is that efforts to second guess the device cannot help you decide what to do.” Korsgaard believes her thought experiment shows that a belief in determinism makes no difference for what one is to do within the first-person standpoint. “In order to do anything, you must simply ignore the fact that you are programmed, and decide what to do—just as if you were free. You will believe that your decision is a sham, but it makes no difference.” In the thought experiment, Korsgaard seeks to show that even if one grants the (compatibilist) operations of the determinism

195 Korsgaard, Creating the Kingdom of Ends, 162.

196 Korsgaard, Creating the Kingdom of Ends, 163.

197 Korsgaard, Creating the Kingdom of Ends, 163; Korsgaard’s emphasis.
device and allows that the person whose thoughts are determined in this way knows that the device is operating, it makes no difference to the exigencies facing that person as she decides what to do. The reflective task of deciding what to do has not been abrogated. Because of a “heautonomous” self-reflexivity that carves out a space of judgments that one self-legislates for oneself alone one can never do other than to act as if one is free and make a decision to do something based on whatever reasons one finds oneself to have—even if those are believed to be determined by external forces. The self-reflexivity performatively forced by the need to decide what to do is thus practically isolated from that which is materialistically explained. A moment of performative inevitability proscribes any vitiating influence from without that would determine that activity of choosing within the practical circle of self-reflexivity.

While Korsgaard’s thought experiment is interesting, it does not do the work she needs it to do. The worry that needs to be addressed is not just a general determinism, but a determinism that accounts for a specific class of someone’s thoughts, namely, thoughts that one “should” or “ought to” do something. Imagine that you have the same device implanted in your brain, but now you’ve been told that it will determine your thoughts so that you decide to sell all your possessions and give the money to the poor. This, let us say, is contrary to what you would otherwise decide to do. Imagine also that you are made aware that the device can be overridden by a determined cultivation of a conflicting desire. Once the device is flipped on and you find yourself having thoughts to sell all your possessions, what will be your relation to those “reasons”? Your belief about the provenance of those “reasons” gives you additional reason to resist them from within your first-person deliberations. You might very well say something like the following as you
think through what to do: “Well, I have these reasons to sell all my possessions and give the money to the poor, but I know why they are there. If it weren’t for this machine that’s running, I wouldn’t be seeing these “reasons” as reasons at all. So, I think I will work to resist these “reasons” as well as I can. In fact, I think it would be a really good idea to make the attempt, since acting on these reasons would be very costly.” The “reasons” within this situation take on more of an optional flavor—or even come to seem like they should be resisted—because of the explanatory story you accept about them. One can easily see how accepting a materialist, event-causal story about moral reasons might direct one’s reflections in similar ways. This modified thought experiment is closer to our real life situation, where the agent is wondering specifically not about a general determinism, but about the status of the moral reasons that seem to be pointing to a costly behavior.

But what if the device is known to be determining both the thoughts that lead you to have reasons for selling all your possessions and giving the money to the poor and the countervailing thoughts about the extreme costliness of the behavior? It’s not clear that this will have no impact on your behavior, as Korsgaard seems to think would be the case. Instead, it seems likely that one might try to adopt—insofar as possible—a sort of Pyrrhonian stance to the whole business and attempt to attain a state of ataraxia. Given your belief about the provenance of the “reasons” you have on all sides of the issue, from the first-person standpoint you might well decide that the best course with respect to all these “reasons” is not to get too excited about any of them and to live a life according to the

198 It does not seem that Korsgaard would have some sense of “overdetermination” by the will in view here, but rather the typical compatibilist view that what a person desires determines her will. Even this, Korsgaard is arguing, does not obviate the standpoint from which one must reflectively decide.
“practical criterion.”

Neither this nor that,” might well be your motto. And this, of course, is a long way from the Kantian moral life of agonistic resistance to “pathological” determination of the will, a resistance that Korsgaard is urging on her readers.

Summing up this first problem for Korsgaard with respect to her transition problem, she cannot grasp this second horn of the dilemma any more than she can grasp the first horn. The second horn cannot be grasped because the third-personal explanatory stories that an agent herself (first-personally) believes do in fact shape the way she will reason within the first-person standpoint when deciding what to do. In fact, when she accepts materialist, event-causal explanations as her own, then she has readily understandable reasons relevant to her decision process in the first-person standpoint for resisting the impositions of moral “reasons”—at least viewing them skeptically, and perhaps actively resisting them in the name of advancing other reasons that may then look all the more persuasive. But if Korsgaard doesn’t find this second horn of the dilemma comfortable, neither can she grasp the first horn and say that moral agents must rely on a practical faith that we really are free (and, Kant himself would say, that we would also have to act as if we have a soul and there is a God). To do so, again, is to fly in the face of the Kantian architectonic she accepts and to say something about how things must actually be ontologically (a problem Kant also faces). But this requires flying in the face of all cognitive explanation, and for those like Korsgaard who do in fact accept the seamless causal explanatory scheme of the Modern Scientific World View (even if events are themselves indeterministic), it also involves a curious division of oneself in which one is constrained to

---

199 Such a person will attempt to be content, after having warded all dogmatisms, with a life of ataraxia according to the “practical criterion” (See Sextus Empiricus, Outlines of Pyrrhonism, 1.23 in Hellenistic Philosophy: Introductory Readings, trans. Brad Inwood and L.P. Gerson (Indianapolis: Hackett Publishing Company, 1988), PH 1.21-24 and PH 1.236-241).
have a faith that one’s own best explanatory account is in fact false in order to be moral. This is not an enviable position either, but these options exhaust the field for Korsgaard (either practical faith in free intelligible causality is or is not needed for morality). Korsgaard’s materialism closes off any avenue by which she might escape this dilemma.

A second, and related, problem tied to Korsgaard’s transition problem has to do with what Bernard Williams spoke of as the “transparency” of moral reasons, and is a central focus in the opening pages of Korsgaard’s *The Sources of Normativity*. Korsgaard herself, in laying out three conditions that any answer to what she calls “the normative question” (the agent’s question within the first-person standpoint regarding what she has reason to do), identifies this notion of “transparency” as the second condition that must be satisfied.\(^{200}\) This is how she puts the matter: “If a theory’s explanation of how morality motivates us essentially depends on the fact that that source or nature of our motives is concealed from us, or that we often act blindly or from habit, then it lacks transparency. The true nature of moral motives must be concealed from the agent’s point of view if those motives are to be efficacious.”\(^{201}\) This is relevant to the question of this section because for Korsgaard, the central point of the reasoning of the agent within the first-person standpoint is that the agent is able to respond to reasons and to be motivated to act by those reasons.\(^{202}\) But Korsgaard’s theory undermines reasons such that they cannot have this capacity so long as one adopts Korsgaard’s moral theory and her materialism. That is to say, her theory

---

\(^{200}\) Korsgaard, *The Sources of Normativity*, 17. The other conditions are the answer must actually address the agent first-personally (must be a reason “of the right kind,” as Stephen Darwall puts it), and that it must “appeal, in a deep way, to our sense of who we are, to our sense of our identity” (17).

\(^{201}\) Korsgaard, *The Sources of Normativity*, 17.

\(^{202}\) This could lead one into a mare’s nest of problems regarding theories of motivation and the like. For the purposes of this section, these issues can be skirted. They have been addressed to some extent in Chapter 2 and they will make an appearance again in Chapter 5.
undermines reasons to the extent that her own moral theory fails to meet the transparency condition for answers to the normative question.

Korsgaard’s difficulty meeting the transparency condition may be seen even more clearly when one imagines oneself in the first-person standpoint of a Korsgaardian agent. How is the Korsgaardian agent within the first-person perspective constrained to view her reasons, and—when viewed in this way—do those reasons maintain their ability to address her first-personally and motivate her by means of this address? In short, what do reasons look like from the perspective of one who accepts Korsgaard’s moral theory and her materialist explanatory scheme? Here again, Korsgaard’s transition problem rears its head. The problem for Korsgaard is that the agent (who accepts an explanation of the reasons she is weighing in the first-person perspective in which those reasons are explained in terms of non-moral causal forces appropriate to the materialism she holds to be true in the third-person standpoint) may now see such reasons as having lost their status as reasons addressing her in her capacity as a deliberating agent. Reasons don’t hold up as reasons that address the agent herself as a moral agent. Because of the materialist explanatory account she accepts, she sees (or at least believes that she sees) that her reasons are really a kind of shorthand for a cluster of causal forces that have her in their grip. Upon coming to see this, she may, as a first-personal agent deciding what to do, no longer treat those reasons as reasons addressed to her and making a claim on what she should do, but rather as the product of causal forces whose grip is to be escaped if possible. In short, once the nature of the forces truly motivating the agent (in the guise of her “reasons”) are seen for what they are, they lose their ability to motivate her as reasons that she sees as offering support for a course of
action and on which she should act.\textsuperscript{203} The authoritative prescriptivity of moral reasons goes missing. Here again, it makes sense from the first-person perspective to regard what she had taken as her “reasons” to have prescriptive authority given the explanations she has accepted. The agent who accepts Korsgaard’s moral theory and her materialism is left with “reasons” that have lost what Williams calls their “transparency.” The support relations she had taken herself to have been appreciating, weighing, and acting on as she justified her actions by referring to their authority now drop from view, as all she sees are causes.

The problem here presses all the harder when, as Korsgaard reminds her readers, “the day will come, for most of us, when what morality commands, obliges, or recommends is hard… And then the question—why?—will press, and rightly so. Why should I be moral?”\textsuperscript{204} But the moral reasons Korsgaard can offer, caught up as they are in the transition problem, do not have the ability to meet this pressing challenge that she rightly raises. It’s not clear why the agent should be motivated to do something costly for moral reasons, when she accepts an explanation of these reasons that subsumes them without remainder in the non-moral, causal terms of the Modern Scientific World View. When those moral reasons—understood as owing to non-moral contingencies regarding one’s evolutionary history, biochemical composition, and such—are pitted against a stiff “why?” question of the sort that worries Korsgaard, it is obvious that their ability to move someone to do what “should” be done will have been weakened by this understanding. Basically, one

\textsuperscript{203} Rue, \textit{By the Grace of Guile}, consists of a wholesale acceptance of the speculations of sociobiology followed by an extensive attempt to preserve morality by means of a “noble lie” that will keep morality intact despite the fact that Rue believes it has lost what Williams would call its “transparency.” Rue thinks it is of utmost importance that this lack of transparency be concealed from the herd, who will not be strong enough to handle it (as he and a few others of the cognoscenti can) and will run amok.

\textsuperscript{204} Korsgaard, \textit{The Sources of Normativity}, 9; emphasis in original.
is left to wonder about the status of this costly “should” when the moral reasons undergirding it are accepted as having a non-moral, materialistically credible explanation, and that the “reasons” and consequently the “should” one feels first personally would be different had certain non-moral events played out somewhat differently in one’s life.

The problem is amplified when one pauses to consider the reasons others give for their moral behavior that “contradicts” one’s own behavior, and remembers that their moral “reasons” share the same footing as one’s own. As one observes the moral behavior of others, one may frequently end up explaining away their putative “reasons” for their behavior by noting how they may be explained in other, non-moral terms. Their “reason” for a particular behavior really owes to their economic self-interest, or their evolutionarily programmed interest in cooperating with the group that is relevant to the success or failure of their self-interested projects (as Gibbard might say). But then one is aware that they could (and probably do) say the same thing about one’s own “reasons.” Moreover, one does in fact accept that there must be some such explanation of one’s own “reasons.” In such a situation, it is difficult to justify to oneself the reasons for insisting that one’s own action is uniquely picked out as being what one “should” do. In addition, in the face of so many competing “reasons”—each having the same epistemic and justificatory footing as one’s own when viewed from the materialist, event-causal explanation that one does in fact believe—it becomes all the more difficult to do what is “right” when that is costly and when

---

205 This simply lays aside the Frege-Geach problem that Korsgaard might need to face along with Gibbard and speaks of “contradiction” for ease of expression.
one believes one’s own reasons are susceptible of being explained away by those who disagree on the same grounds that one uses to explain away their reasons.\[206\]

In summary, this second problem relating to Korsgaard’s transition problem arises from the fact that when someone accepts a third-personal explanation that subsumes moral reasons under non-moral, event-causal terms of the sort acceptable within the Modern Scientific World View, this explanation—having been accepted by the agent—cannot be isolated from the first-personal standpoint where the agent must decide what to do. When such an explanation is believed by the agent, the explanation ends up explaining away the reason and the moral theory that rests on such reasons ends up failing the “transparency” condition. Here again one is left with a severe problem of moral reasons arbitrariness.

A third and final difficulty related to Korsgaard’s transition problem may be seen by considering a possible objection to the second problem. Perhaps Korsgaard can defuse the second problem by emphasizing that it is the agent herself who underwrites her own reasons, either through her practical identity as a human being or else through the very form of the maxim that itself expresses the distance from the domination of a seamless efficient causality that itself constitutes morality. Perhaps, to use the Kantian language, there is a sort of heautonomy that can save moral reasons by showing their inherent self-reflexivity.\[207\] It is not that there are reasons that the agent somehow stands apart from and explains, but rather she gives herself her own reasons as a product of the very self-reflexivity (either of the

\[206\] In *The Last Word*, Nagel attempts to defuse possible problems a view like his or Korsgaard’s might face in connection with its uneasy relationship to the materialist explanations that they accept (or kind of accept, in Nagel’s case; it’s hard to tell what to make of “panpsychism”). This argument has been addressed in Chapter 2 near the end of the second section.

\[207\] For an account of heautonomy as it relates to Kant’s thought, see pp. 329ff. above.
practical identity or of the maxim) that is constitutive of the first-person standpoint. One cannot, then, stand apart from one’s own reasons in such a way as to explain them in terms of a third-personal, materialist explanation, since their very self-reflexive nature does not allow them to be scrutinized and explained in this way. They simply are part and parcel of the first-person standpoint itself.

But there are severe problems facing an attempt to appeal to heautonomy in this way in an attempt to solve the transition problem. The first of these is that the materialist explanatory account Korsgaard accepts can cast doubt on this self-reflexive awareness of one’s own responsiveness to reasons into question as being nothing more than a bit of “folk psychology.” The materialist, event-causal explanation ends up explaining this self-reflexivity in its own terms and both the “clearing” in which reasons come to light and the reasons themselves are explained in ways that call them into question. Allan Gibbard puts the problem well in speaking about the business of giving explanations in relation to the Galilean Core of science: “Again, the question is how we explain our accepting these explanations. Can we do so without citing the kinds of facts that figure in them? Perhaps we might, but if we did, we would have debunked these explanations. We must explain our Galilean judgments as responsive to the kinds of facts that figure in them, or we lose our grounds for making the judgments.” But this is exactly what the Modern Scientific World View claims must somehow be done with respect to human self-reflexivity and the reasons we thereby recognize; they are stripped of their own terms and explained in terms that end

---


up explaining them away. And Korsgaard, as a materialist, accepts that some such explanation does in fact exist. Appeals to heautonomy do not succeed in isolating self-reflexivity and its reasons from the explanations that the Korsgaardian agent accepts as well-founded, and the failure to do so casts such an agent back into the first two problems discussed above.

As seen above, when faced with such challenges, Korsgaard issues statements enjoining her readers who might doubt the reality of colors to look at some paintings, or who might doubt the reality of reasons and values to go make a choice. Upon doing so, they will change their minds.\textsuperscript{210} Undoubtedly so. But then how is this new mood to be explained? Does one feel this way because there actually are reasons that do in fact exist? In which case, where does this leave the materialist explanations that one accepts from within the third-person standpoint? Or does one feel this way owing to causes that must—somehow—fit within the Modern Scientific World View Korsgaard accepts? In which case, again, where does that leave her reasons, since she accepts a view that explains them without remainder in terms that, as Gibbard rightly notes, explain them \textit{away}?

But even if these problems are set to the side, another problem besets the appeal to heautonomy. Having isolated self-reflexivity—whether in practical identity or in the structure of the maxim (the \textit{logos})—from all input except itself, how can the ends it gives itself in adopting an identity or following a maxim be reined in so that the self-reflexivity is constrained to go in specifically moral directions rather than in any direction whatsoever? Isn’t this isolated self-reflexivity free to set any end at the heart of its practical identity or maxim? There seems to be a problem similar to that which faced Gibbard with respect to

\textsuperscript{210} See note 19 above.
setting the ends in plans—a problem Gibbard failed to solve.\textsuperscript{211} Like Gibbard, Korsgaard fails to solve the problem of fixing ends. Their failure should be worrisome to materialists.

The worry, to which both Gibbard and Korsgaard are responding in their different ways, is quite simple. As Kant put the matter, it is that morality might turn out to be a mere “chimerical idea.”\textsuperscript{212} Korsgaard puts her finger on the nub of the problem: “A theory that could explain why someone does the right thing—in a way that is adequate from a third-person perspective—could nevertheless fail to justify the action from the agent’s own, first-person perspective, and so fail to support its normative claims.”\textsuperscript{213} Materialists accept an ontology that disallows personhood in the fundamental explanatory terms tied to that ontology.\textsuperscript{214} But then where does this leave the irreducibly personal nature of a person’s first-person experiences and deliberations about what to do? This is the heart of the transition problem that so worried Kant—a problem that in different ways exercises Korsgaard and Gibbard in turn.\textsuperscript{215} What, if anything, can be salvaged of the authoritative

\textsuperscript{211} Korsgaard on maxims is in many ways strikingly parallel to Gibbard on plans. As Gibbard notes, there are deep internal similarities to their projects—at least in their starting point in reflective endorsement.

\textsuperscript{212} Kant, \textit{Groundwork of the Metaphysics of Morals}, 51/4:445. Darwall quotes this passage several times in \textit{The Second-Person Standpoint}, evidencing that he is quite worried about this.

\textsuperscript{213} Korsgaard, \textit{The Sources of Normativity}, 14.

\textsuperscript{214} I confess to being more or less “underwhelmed” by so-called “non-reductive” materialism. If, at the beginning, there was only non-personal stuff of whatever type materialists are willing to countenance, then there must be some explanatory route from that stuff to what there is now. It seems to me that materialists are committed to at least some sort of token-reductionism. Jaegwon Kim is right to inveigh against having “Physicalism on the cheap,” (Kim, \textit{Mind in a Physical World}, 120). As David Papineau points out, however, in defending what he calls “token congruence,” an ontological requirement of some sort of (what I am calling) a token reduction, does not mean that “the practitioners of the special sciences have to know that reduction,” (David Papineau, \textit{Philosophical Naturalism} (Oxford: Blackwell, 1993), 40; Papineau elaborates on “token congruence on p. 12). I have addressed these issues, including possible ontological ramifications of this “explanatory gap” (as Joseph Levine has called it), more fully in Seeman, “Out-Kimming Kim,” and I cannot elaborate on them here.

\textsuperscript{215} See note 161 above.
prescriptivity of moral reasons given materialism’s drive toward third-person, event-causal explanation that has little, if any room for a non-reduced normativity?

Faced with this thorny problem, Gibbard turned outward in attempting to find a ground for moral reasons (or something close enough), an attempt that has been shown to end in failure because the stuff Gibbard finds in the “Galilean Core” of materialist explanations simply cannot support the needed normativity. Gibbard finds plenty of external causal forces, but never finds reasons or anything particularly close to them, with things like the authority of reasons being glossed as “conversational pressure” (which of course cuts across all human moral distinctions as something anyone pumping any agenda can exercise on anybody).

Korsgaard, on the other hand, sees the inadequacy of non-personal, materialist, third-personal explanations to vouchsafe even a facsimile of first-personal reasons. Thus, in a quintessentially Kantian move, she looks inward. In this she follows Kant’s strategy of fitting the first and third person standpoints together through heautonomy, that is, a fundamental self-reflexivity in which there is not only a judgment, but the constitution of the domain over which the judgment ranges—all spinning in isolation from the third-personal realm of non-personal, materialist explanation. And so one finds Kant, speaking (not coincidentally) about purposes or ends in our cognition of nature, saying that these laws “must be considered in terms of the sort of unity they would have if an understanding (even if not ours) had likewise given them for the sake of our faculty of cognition, in order to make possible a system of experience in accordance with particular laws of nature. Not as if in this way such an understanding must really be assumed… rather this faculty thereby gives
a law only to itself, and not to nature.”

The “as if” of self-reflexivity dominates in Kant, and Korsgaard follows this pattern.

The fundamental problem for materialists who attempt to turn inward flows from the nature of the attempt: one is turned in on oneself, with no normative limits upon that self from without. If the normative constraints do not come from without (since the materialist stuff outside human beings is not able to provide it), the question is how the needed constraints will be manufactured within one part of that materialist world: homo sapiens. When the materialist turns inward, it’s not clear how what she finds within herself will be any more capable of providing normative constraint than what exists outside herself. At least she might find prescriptivity (even if she can’t explain it in materialist terms), but the attempt to locate some sort of authority for such prescriptivity is highly problematic. I have argued that Korsgaard’s attempted bridge connecting the first-person standpoint of personal reflection and decision based on reasons to the third-person standpoint of causal explanation in non-personal terms fails to cross the “chasm” Kant saw. The first-person realities do not extend beyond oneself to make a normative claim on another, nor to force the other to be taken into account normatively rather than merely as a factor to be considered in calculations of what is “viable.” I have argued that even if the subject can value her own reasons and act upon them herself (which I don’t think Korsgaard has established), there is no need for her to value someone else’s reasons as reasons for her to act. Nor do her reasons for acting extend beyond herself to give someone else a reason to act. In attempting to hermetically seal the first-person standpoint from the vitiating causal forces of the third-person standpoint, the first-person standpoint is sealed off from everything.

---

216 Kant, *Critique of Judgment*, 67–8 (5:180)
outside itself. Indeed, for the self-reflexive subject, the claimed first-person standpoint of other people is not available and is liable to complete non-personal explanation in third-personal causal terms. Unlike her own experience of her own first-personal standpoint, there is little, if anything, in her experience of particular primates outside herself that must be understood in ways that disrupt the complete causal explanation of their behavior. When considering their behavior, the first-personal standpoint that she cannot avoid in her own reflections and that thus keeps her from closing herself in the system of non-personal causal explanation simply is not there to disrupt the causal closure. After all, she does not experience their first personal stance, and in Korsgaard’s self-reflexive moral philosophy it is this that forces the deciding subject into the stance in which she cannot do other than decide how she will act in the world. The problem for the homo sapiens apart from oneself in Korsgaard’s moral philosophy is that Korsgaard’s insistence that “you must choose as if you were free” does not mean that you must choose as if they were free. They may be fully explained, and need not be treated as ends in themselves or have their reasons taken into account as equal partners to be consistently harmonized. Again, that which supposedly prevents you from disregarding your own humanity—the practical necessity of deciding what you yourself will do—is not something that you have access to in the case of homo sapiens outside yourself. Any action the subject takes is consistently universalized over all first-personal subjects she knows—but at best she knows only of herself in the requisite way.

Not only does Korsgaard’s self fail to break out of the ambit of self-reflexivity, it is not clear how even this one subject fits with the causal explanation the materialist herself accepts. As Gibbard was seen to point out (rightly) above, when something is explained in

217 Korsgaard, Creating the Kingdom of Ends, 162; emphasis added. See pp. 311-13 and 349-51 above for some criticisms of “universalizability” arguments.
terms fundamentally different from the terms fundamental to its own patterns of explanation, it is effectively explained away. Yet the materialist is committed to the existence of some non-personal explanation of everything else—including personhood. After all, the non-personal stuff, however that is understood, came first. This is a basic commitment of materialism; it is the whole point, really. Non-personal stuff came before persons, and persons must (somehow) be explicable in terms of the prior non-personal stuff, since, ontologically, it must have happened some way or other.

Having accepted the existence in principle of some such explanation, even if she has no idea how that could go, the materialist casts a jaded eye on the authoritative prescriptivity of even her own moral reasons. Eventually, as Gibbard sees, the giving of replies to “Why?” questions eventually comes to rest in a substantive starting point. And for the materialist this substantive “why stopper” will reach ontological bedrock in the non-personal, material stuff. And both the options Korsgaard explores for overcoming this limitation—identity-priority maxims and form-priority maxims—founder at precisely the point where they run into this non-personal ontological bedrock.

Between them, Korsgaard and Gibbard explore two fundamental options for materialists who would not simply (and implausibly) eliminate authoritatively prescriptive moral reasons altogether: a turn inward to a fundamental self-reflexivity or a turn outward to a third personal facsimile of the first personal carried out by means of non-personal causal forces. Neither of them succeeds in giving a satisfactory account of moral reasons.

---

218 As noted in the Introduction, the inward turn/outward turn language is not meant to specify an exhaustive set of options. Although I do in fact think that in some broad sense these options are exhaustive, I need not argue for that here and I make no attempt to do so. I welcome readers to think through the matter themselves.
Turning outward to the non-personal causal realm of materialism, Gibbard finally fails to overcome the limits of external, non-personal *causes* that he (rightly) sees himself as limited to in explaining morality. Normativity in general—and moral reasons with it—goes missing. In place of moral reasons, Gibbard gives only external *causes*. First-personal moral reasons thus come apart from any constraint that is itself moral (or so I argued in Chapter 3). Turning inward to a realm of first personal self-reflexivity, Korsgaard finally fails to connect these *internal* reasons to the third-personal causal explanation of the world that she, as a materialist, *accepts*. There is no sense of how any of this could have happened, given a materialist world.\(^{219}\) And this in a twofold manner: she neither shows (1) how a solipsistic self-reflexivity could take the “reasons” of others (who are themselves explained without remainder in non-personal terms, and whose own first-personal experience is not available outside themselves) into its own first-personal stance so as to be required to take their reasons as one’s own; nor (2) how any of this could fit into the non-personal causal story of third-personal explanation she accepts—thus showing it to be something more than a “figment.” On the one hand there is the inward turn to a self-reflexive first-personal standpoint hermetically sealed from all the third-personal forces that would vitiate it. Or, on the other hand, there is the outward turn to external, non-personal causal forces that somehow provide a workable facsimile of moral reasons.

In the hands of Korsgaard and Gibbard, both the inward turn and the outward turn fail. There are, of course, many different, creative attempts to pursue one or the other of these paths toward a workable materialist account of moral reasons, and in Chapter 5 John

\(^{219}\) After all, if materialism is true, the non-personal, non-normative material stuff came before all else, and anything that cannot plausibly be thought to have arisen from that material stuff cannot be true if materialism is true. But Korsgaard has given no plausible story of how agents or maxims or standpoints or other parts of her Kantian machinery could have arisen from the material stuff.
Post’s attempt to locate a normativity that arises from outside forces that give rise to human beings will be examined. I will argue that Post’s attempts do not succeed any more than those of Korsgaard and Gibbard. And though this dissertation will not argue that all such attempts must end in failure, one might wonder if the reasons for the failure of their attempts seem to owe to something more than mere idiosyncrasies of their ways of trying to preserve something of authoritatively prescriptive moral reasons within materialism.

IV. Korsgaard and Darwall

Having tried to point out some difficulties with Korsgaard’s moral philosophy, this chapter will conclude by comparing Korsgaard’s first-person approach to moral philosophy with Stephen Darwall’s approach, which at first appears to be a second-person approach to morality. Yet Darwall is a Kantian, and as such he finally also brings the second-person approach under the thumb of the first-person standpoint. That said, it will be seen that both Korsgaard and Darwall have some important insights into the nature of moral reasons.

One important insight emerges in connection with Korsgaard’s attack on consequentialism, where Korsgaard states flatly that consequentialism “in a certain way is not actually a moral theory… It is a kind of technological vision of something you would put in place of a moral theory. It is a social engineering project.” What Korsgaard is assailing is a problem that consequentialist theories have with moral reasons; the “reasons” on offer turn out not to be reasons of the right kind. As she puts it in The Sources of

---

220 Korsgaard, “Internalism and the Sources of Normativity,” 57. Note that I don’t intend this as an argument against consequentialism generally (something that would be outside the scope of this dissertation). Instead, I am taking this to raise an important insight about reasons. If Korsgaard is right that consequentialism generally fails to provide “reasons of the right kind,” then I would take it that consequentialism might be in trouble. But I take no stance on that here.
Normativity, “Mill has lost track of an essential point. The normative question must be answered in a way that addresses the *agent* who asks it.”\(^{221}\) The agent must be given a reason to which he can respond as an agent, rather than merely being determined in his behavior by calculations of how to enter effectively into a world explained in third-personal terms.

Korsgaard’s solution is a Kantian solution: The agent addresses reasons to himself. A personal moral agent responds to moral reasons that address him as a moral agent because they have their source in a moral agent, namely, himself. Korsgaard’s solution, then, is a first-person solution. Moral reasons are first-personal: “The normative word ‘reason’ refers to a kind of reflective success.”\(^{222}\) This is a consistent theme in Korsgaard’s work, as has been seen above. A moral reason is what a person has at the end of a process of reflection as an autonomous agent, and it addresses the person as a moral agent because it is something that emerges out of his very identity as an autonomous agent (or in the maxim-structure or *logos*) in addressing a law to himself. The ability to receive moral reasons is called out in the very act of reflectively addressing a law to oneself. To be a moral agent is just to have given oneself the right kind of reason in freely choosing what to do. In addressing oneself freely one cannot help but have the right kind of reason for action; as a moral person you are addressing a reason to yourself. And this is the important point: a person addresses reasons to a person.

That reasons are something that only persons can have and only persons can give is an important insight, and Kantians are exactly right in clinging to this. This is a requirement of “reasons of the right kind.” Some problems with Korsgaard’s attempts to save such

\(^{221}\) Korsgaard, *The Sources of Normativity*, 85-6; emphasis in original.

\(^{222}\) Korsgaard, *The Sources of Normativity*, 93.
reasons from the materialist acid of non-personal third-person explanations have been
explored above, and they will not be repeated here. Instead, here it will be noted that a
divergent strand in Korsgaard’s thought about moral reasons holds promise for moral
reasons that are of the right kind while avoiding the problems plaguing Korsgaard’s first-
personal account of moral reasons. Korsgaard concludes her Tanner lectures with the
following words before interacting with her interlocutors: “it is the most familiar fact of
human life that the world contains entities that can tell us what to do and make us do it.
They are people, and the other animals.”

There is an incipient divergence from the first-
person standpoint that is never developed in Korsgaard’s work, although it also appears
elsewhere: “To say that you have a reason is to say something relational, something which
implies the existence of another, at least another self. It announces that you have a claim on
that other, or acknowledges her claim on you. For normative claims are not the claims of a
metaphysical world of values upon us: they are claims we make on ourselves and each
other.”

Perhaps reasons of the right kind to address us as moral agents can be given
second-personally, so that reasons are fundamentally relational. Perhaps another might have a
rightful claim on us. Another person addresses an authoritative claim to us and we must
respond to the reason that has in fact been given to us.

This does not immediately seem to hold promise as a way of avoiding moral reasons
arbitrariness, for there will be as many different reasons addressed to us as there are people,
and these will undoubtedly diverge. Somehow these reasons addressed to others must

---

223 Korsgaard, *The Sources of Normativity*, 166.

224 Korsgaard, *Creating the Kingdom of Ends*, 301; emphasis on the words “claim” and “claims” is added. See also, 205.
themselves be constrained, and this is exactly what the first-person standpoint is supposed to do in Korsgaard’s thought. It is of the essence of a Kantian moral philosophy to award the palm to the first-person reasoning of the autonomous agent when faced with the question of authority in the moral life. In both the passages immediately above it is clear that the second-person address is not fundamental for Korsgaard’s thinking about moral reasons. First, the “other animals” Korsgaard mentions in the first passage are not persons and do not address reasons to us; they might perhaps be the occasions of reasons arising in us as we think autonomously about what to do. There are no second-personal reasons here, but only first-personal reasons that take the animal into account. Second, the second passage above leads into the concluding line of the essay: “the only reasons that are possible are the reasons we can share.” This may sound second-personal, but it is not. Rather, other autonomous agents will find themselves sharing our reasons as they, too, reason autonomously about what to do. What is fundamental is not the second-person address but the first-personal reasoning that carries both agents toward the same moral reasons that they each address first-personally to themselves. The second-person claim must first pass through the first-person standpoint before it becomes a moral reason; moral reasons are always something moral agents give themselves first-personally. Jean-Jacques Rousseau put the central idea nicely when he formulated an ideal of political association among people “by means of which each, uniting with all, nevertheless obeys only himself.”

---

225 Korsgaard, Creating the Kingdom of Ends, 301.

standpoint is swallowed up in the first-person standpoint, with the latter finally being swallowed up in a third-personal standpoint that finally requires all fundamental explanation to be grounded in the postulated non-personal stuff that ontologically is the root of everything. In Korsgaard the second-person standpoint is an aside.

Although Stephen Darwall goes to great lengths to develop a second-personal standpoint in morality, as a Kantian he finally retreats to the same first-person standpoint that grounds moral reasons in Korsgaard’s thought. It is revealing that the passage above from Rousseau is quoted in Darwall’s book, *The Second-Person Standpoint*, as he works out a Kantian contractualism. Darwall, as was seen at the end of Chapter 3, works out a moral philosophy that attempts to give the second-personal address and the authority of the second-person a fundamental place in grounding moral reasons. In the course of working out this grounding, Darwall develops some important ideas: the second-person address, as well as the summons and authority of the second person. At the center of the second-personal account of morality Darwall starts to develop are what he calls second-personal reasons. Darwall is clear that these are normative reasons, and when these reasons are addressed to another person, they are “there anyway, independently of your getting him to see it or even of your ability to do so.”227 As seen in Chapter 3, this kind of independence is a key feature of moral reasons, and in Darwall’s thought it initially seems to be grounded in second-personal authority: “Our dignity as persons includes,” Darwall maintains, “an irreducibly second personal authority to demand respect for this very authority and for the

---

requirements with which it gives us the standing to demand compliance.”228 An authority inherent in the second-personal address seems to ground second-personal reasons. These are reasons of the right kind, addressed by a person to another person as a moral agent.

But what will limit these reasons, so that not just any reason can be offered by any agent as an authoritative reason for action? Here Darwall’s Kantianism comes to the fore. Second-person reasons have no voice of their own apart from the authority of the first-person standpoint; a second-person reason is only a reason insofar as it is reaffirmed in the first-person standpoint by the autonomous agent.229 The name Darwall settles on for his core idea—“morality as equal accountability”—is revealing. Darwall argues that divine-command theorists backed their way closer and closer to a position that is summed up in what Darwall calls “Pufendorf’s Point”: “in holding people responsible, we are committed to the assumption that they can hold themselves responsible by self-addressed demands from a perspective that we and they share.”230 Even God, these theorists concluded (at least on Darwall’s reading), could not hold other agents responsible unless they themselves could freely ratify the reasons for complying. But if this is true, then it is a short step to recognizing that we can also hold ourselves accountable on our own, and God becomes morally dispensable. Darwall goes ahead and dispenses with Him and defends this morality

228 Darwall, The Second-Person Standpoint, 14.

229 Darwall recognizes, or rather insists, that his view is finally a first-personal view: “Since second-personal reasons are always fundamentally agent-relative, the second-person stance is a version of the first-person standpoint (whether singular or plural).” (Darwall, The Second-Person Standpoint, 9).

230 Darwall, The Second-Person Standpoint, 112; this entire passage is italicized in the original. It’s possible that Pufendorf might have meant that this holding of oneself responsible was something possible in principle. I am not sure that even this would be wholly unproblematic, but the point of the story Darwall is telling about Pufendorf is that it opened the way to philosophers like Kant who definitely did not think that human beings did this only in principle.
as equal accountability—a fundamentally Kantian picture that makes the first-person standpoint of the autonomous agent fundamental to all moral reasons: “any second-personal address whatsoever presupposes an addressee’s capacity to accept and act on reasons that are grounded, ultimately, in an authority that addresser and addressee share as free and rational.”231 Second-personal reasons thus have no standing whatsoever apart from the first-personal standpoint of the free and rational agent. But this first-personal standpoint must now be defended as fundamental, and nothing of all the machinery Darwall develops around the second-person standpoint amounts to anything unless this more fundamental project can be carried forward. But here Darwall faces the same kinds of problems that I have argued that Korsgaard faces.

Korsgaard and Darwall clearly see that moral reasons must be reasons of the right kind to address moral agents as persons, and both haltingly take steps toward seeing the relevance of authoritative second-personal claims that are addressed by one person to another—although neither can ultimately find any place for second-personal reasons to have a voice apart from the reasons one gives to oneself. The second-personal standpoint collapses into the first-personal standpoint for both Korsgaard and Darwall. Indeed, it is very unclear what hope there could be for solving the fundamental problem of second-person moral reasons from within the Modern Scientific World View. Such reasons could only come from this person and that person, and unless there is some more basic check on what will count as a reason, there will be a welter of conflicting reasons offered up. That is, a moral theory that weds the second-personal standpoint to materialism collapses into moral reasons arbitrariness.

Perhaps a key to the problem is found in a slogan Darwall repeats frequently in *The Second-Person Standpoint*: “Second-personal reasons out, second-personal reasons in.”

Though I do not argue the point in this dissertation, perhaps if a moral philosophy starts with only the sorts of stuff countenanced by the Galilean core of the Modern Scientific World View that Darwall, Korsgaard, Gibbard, Post and many others accept, there will be no way to take the non-personal starting point and transform it into the second-personal input that is needed. You might get a stability in external forces that fails to provide reasons of the right kind, or reasons of the right kind arising from the unstable endorsement of this individual or that individual (or group of individuals) that gives reasons of the right kind, but that vary wildly. Either way, one would have moral reason arbitrariness, because the “second-person in” is highly unstable (think of the welter of such inputs) apart from second-personal reasons addressed to all people by an authoritative person and the proposed substitutes fail to provided the needed independent, authoritative prescriptivity.

But what if the second-person standpoint were fundamental—such that third-person explanation ultimately came to an end in a person addressing all other persons? Then reasons could be fundamental and external in a way they could not be otherwise. This possibility will not be explored in this dissertation, though the Conclusion will suggest that the argument of the dissertation as a whole opens out into an exploration of the possibility of an ethic that is second-personal all the way down.

Be that as it may (or may not) be, I have argued that Korsgaard’s attempt to underwrite authoritatively prescriptive moral reasons fails. In addition to real questions about how her moral philosophy squares with her adherence to the Modern Scientific World View, Korsgaard’s moral philosophy faces the circle of the Identity/Reasons
problem. Whether Korsgaard privileges identity over reasons in what I have called the
identity-priority maxims view, or whether she begins with reasons before identities in the
form-priority maxims view, Korsgaard does not finally avoid moral reasons arbitrariness.

It was also suggested that Korsgaard has a real question of how the morality of the
first-person standpoint fits with the materialism of third-person explanations, and that
Korsgaard’s moral philosophy attempts to deal with this problem by means of a
fundamental self-reflexivity that is not unrelated to Kant’s notion of “heautonomy.” While
this connection is not needed for the main argument of the chapter to go through, if the
connection is successfully drawn it might point to a deeper flaw affecting Korsgaard, and
possibly other contemporary Kantian moral philosophers as well. When such philosophers
attempt to isolate the distinctly human business of appreciating, weighing, appealing to, and
acting on moral reasons (a business Kantians rightly see as essential to morality) from the
materialism operative in the explanatory realm by appealing to a kind of self-reflexivity, they
end up not only insulating human activities of deciding what to do from materialism, but
also cutting off such individual, self-reflexive activity from authoritative claims addressed to
them from without. Or, to put it another way, the materialist accepts that everything—
including homo sapiens and their behavior—may be explained materialistically, but finds
that she cannot herself help but decide as if free. The problem is that, having thus
(supposedly) self-reflexively isolated the first-person standpoint from that which is explained
materialistically, her self-reflexive activity of deciding what to do is thereby isolated from
those organisms; since other homo sapiens are explained materialistically, they do not have
to be regarded in the same way one cannot help but self-reflexively view oneself. No
amount of consistency will solve this problem for any Kantian who at some point appeals to
self-reflexivity, because that very self-reflexivity excludes everything outside of the self-reflexive circle from having a justificatory voice within that circle. From the self-reflexive standpoint of deciding what to do, the self need be consistent with nothing that is not itself—and the very thing that (putatively) forces the deciding self to recognize itself as “human” finds no similar performative inevitability that forces her regard anything else as “human” in the same way. Any homo sapiens outside the circle of self-reflexivity may consistently be left out of account as she confers value to her ends.

The deeper Kantian connection is not needed to display Korsgaard’s own problems avoiding moral reasons arbitrariness. If my arguments are right, Korsgaard’s attempt to craft a normatively binding materialist moral philosophy by turning inward meets with no more success than Gibbard’s turn outward. In the final chapter of this dissertation, I argue that John Post’s attempt to press Ruth Garrett Millikan’s materialist account of language and intentionality into service of moral philosophy also fails to avoid moral reasons arbitrariness. Like Gibbard, Post turns outward to the material world in an attempt to find the normativity we human beings do in fact respond to in our moral decisions. But like Gibbard, Post fails to capture a meaningful sense of what Darwall calls “second-person authority,” and moral reasons are thus finally undercut. As we have seen Darwall put it, “second-person authority out, second-person authority in.” Neither Post nor Gibbard see their shared materialism allowing them to understand second-person authority to be fundamental. Rather, it is non-personal, material stuff that comes before everything else. Once again, the independent, authoritative prescriptivity characteristic of moral reasons goes missing as Post finally finds

---

232 I take this move to be of the essence of Kantian moral philosophy generally (or at least all materialist versions), but I am not arguing that point here.
only causes where he hoped for the genuine normativity of moral reasons. Like both Gibbard and Korsgaard, Post fails to avoid moral reasons arbitrariness—or so I will argue.
The title of John Post’s most recent book, *From Nature to Norm: An Essay in the Metaphysics of Morals*, sums up a philosophical project that he—and an increasing number of other philosophers—believe must be accomplished if normativity is to be shown to be more than a chimera. Some philosophers see something like this project as necessary to vouchsafe normativity, but see no hope of its completion. Post’s counsel, however, is not one of despair. He believes recent developments in philosophy of language and philosophy of mind, specifically the “teleosemantic” program of Ruth Garrett Millikan and David Papineau, provide the resources needed to show that normativity can arise from the natural world. Specifically, Post thinks Millikan and Papineau have shown how an etiological evolutionary approach to “proper functions” gives rise to biological purposes—a stock example is the heart’s being for pumping blood owing to its evolutionary history—and that, owing to these proper functions, a kind of normativity is embedded in our current best biology. Post thus urges us to “press from below” and “start with something much simpler and more tractable than moral normativity, in order to challenge certain orthodoxies about the place of normativity in the world, [and] then see what implications, if any, this approach  

---

might have for other kinds of normativity.”

As will be seen, Post takes the implications to be quite substantial.

As Post himself notes, however, one of the philosophers whose work he relies on is not at all sanguine about the prospects for grounding any sort of robust moral normativity in teleosemantics. Papineau states flatly: “Wherever the normativity of content comes from, it can’t be from biology, since biology deals in facts, not prescriptions... It has always mystified me why anybody should think that biology helps with normativity.” The worry, Papineau notes, is a divide between what the biological “norms” would be and the moral reasons we take ourselves to have: “My knuckles have arguably been biologically designed to hit people with, but it doesn’t in any sense follow that I ought so to use them. Again, a number of human male traits have undoubtedly been designed to foster sexual predatoriness, but it doesn’t follow that men ought to be sexually predatory.” Or, as G.E.M. Anscombe put it: “The search for ‘norms’ might lead someone to look for laws of nature, as if the universe were a legislator; but in the present day this is not likely to lead to good results: it might lead one to eat the weaker according to the laws of nature, but would hardly lead anyone nowadays to notions of justice.” In short, we seem to have moral reasons that cut across the “norms” that might arise from our evolutionary history, and it is

---


precisely the status of these moral reasons *as moral reasons* that biology doesn’t explain. Or, to put the point in terms of this dissertation’s central argument, the biological facts in which Post attempts to ground the metaphysics of morals afford us external *causes*, but no external moral *reasons*, leaving his project mired in moral reasons arbitrariness.

Throughout this chapter, the distinction between reasons and causes will be in the background, showing up, for example, when a key phrase like “diachronic causal skein” is used to capture an important aspect of Millikan’s theory. The reasons/causes distinction was drawn at length in Chapter 2, and it was noted that there are many difficulties in this area, owing to the fact that philosophers have not succeeded in making either the notion of a reason or the notion of a cause perspicuous.⁶ That said, there are some clear examples, and some key aspects of reasons and causes can be identified. It will be helpful to review some of the main ideas here.

A central aspect of reasons as opposed to causes is that, as Paul Moser has put it, “support relations” are in view.⁷ When a person acts *on a reason*, she acts in view of some support relation between one part of her situation and another part of it. X “leads toward” or “grounds” or “gives a basis for” or “lends support to” Y. These support relations do not “bypass” the agent, but act through the agent’s *recognition* of the salient features of the situation. As Korsgaard helpfully summarizes the matter, the agent “is aware of the reason as a reason; she identifies the good-making properties of the action under the description ‘good’ or ‘reason’ or ‘right,’ or some such normative description. She does not act merely in

---

⁶ See pp. 26-62 above.

⁷ See Moser, *Philosophy after Objectivity*, 160: “An item (for example, event, state, or claim) *X* is a reason for another item *Y* if and only if *X* is a ground, basis, means, or source of support for *Y*.”
accordance with a normative consideration but on one.”

The importance of these features of reasons are clearly in view in the example of Jim’s leg jerking upward when a doctor taps his knee with a mallet (see Chapter 2). There are no support relations that lead Jim to jerk his leg upward, such that he weighs them up and decides, “I think I’ll jerk my leg upward now.” There are no considerations in view at all in this situation. The causal forces bypass Jim’s agency and simply make an action happen.

The situation is quite different with Jane working on her geometry problem and Ted deliberating whether or not to leave a note on the windshield of a car that was damaged when a strong wind pulled his car door out of his hand. Support relations obviously figure prominently in these situations, and Jane and Ted actively have them in view and attempt to weigh them up in an effort to decide what ought to be done or should be done in the situation. They ask: “What do I have reason to do?” And they actively engage in an attempt to identify or discover those reasons and to act upon them. As agents, they are alive to the salient normative features of their situation, and act with a view to identifying those features and acting upon them. As Joseph Raz has put the matter, agents undertake their action and explain them in light of “a story which shows what about the situation or action make it, the action, an intelligible object of choice for the agent, given who he is and how he saw things at the time.” There is an effort (more or less diligent) to identify what renders an object of choice intelligible, and, ideally, the right thing to do in the situation. As we consider these matters, the obvious differences between the situations of Jane and Ted, on the one hand,

---

8 Korsgaard, “Acting for a Reason,” 214; emphasis in original.

9 Joseph Raz, “Agency, Reason, and the Good,” 24; quoted in Setiya, Reasons without Rationalism, 62. Setiya, it should be noted, is advancing a causal notion of reasons that is set against Raz’s understanding of reasons and, indeed, any normative notion of reasons that is operating “under the guise of the good.”
and Jim, on the other, throw into bold relief some essential differences between reasons and
causes regarding support relations and the role of the agent. The agent is alive to features of
the situation that prescribe a course of action as the right thing to do.

Jane and Ted both act on reasons, but their situations also differ. Upon seeing what
he has reason to do, Ted may rebel against those reasons in a way that Jane will not be
tempted to do. Geometrical akrasia makes no sense; but weakness of the will—even more
(and putting things bluntly), disobedience or rebellion—these are features of situations like Ted’s
that are widely, if not universally, familiar. Moral reasons are marked by their authority that
stands over against a desire not to do what we think we have reason to do. As Jean
Hampton observes, “we are moved to act from moral reasons because we understand their
objective authority.”10 Or, as Paul Moser puts it, we are confronted by “Love’s Demand,”
recognizing there a rightful authority over us as persons.11 Both the authority and the
prescriptivity of moral reasons stand in marked contrast to the causes operative in situations like
what happens to Jim’s leg when the doctor taps his knee with a mallet.

Accommodating moral reasons is a tall order for a materialist.12 In Chapter 2 we
saw Dretske struggling without obvious success to make some sense of the address
embodied in moral reasons. Bittner simply dismissed the authority and the prescriptive
address found in reasons as incompatible with his materialism, while Scanlon—with more
sense, but perhaps less consistency—simply took reasons as “primitive” while waving his
hand over the devilish details Dretske labors to work out. I have argued extensively that

10 Hampton, Authority of Reason, 105; emphasis in original.
12 Note 17 in the Introduction gives my reasons for using the term “materialist” instead of
“naturalist.”
these materialist difficulties with moral reasons are clearly on display in the moral philosophies of Allan Gibbard and Christine Korsgaard, and that neither philosopher successfully avoids the problem of moral reasons arbitrariness. The move from nature to normativity is not an easy one. Post holds that the difficulties facing the attempt to move from nature to norm to authoritative moral reason can be overcome, but it is an “arduous climb.”

Direct inference from biological norm to moral norm does not succeed: “we must resist being lured into the futile strategy of inferring a would-be moral ought directly from the biological ought.” Millikan’s “bionorms” cannot be taken over directly into the moral realm. But while Post sees that the attempt to infer “authoritative moral norms from bare facts about nature” is hopeless and even “dangerous,” he believes there is another way forward, one in which biology can show how the epistemic status of our moral norms is justified. Key to this indirect route is the way in which Millikan sees “biological norms” arising from etiology of natural selection, and Post’s arguments will only be as good as the foundations Millikan has laid and will have to work within the limitations of that work—or else give convincing arguments why those limitations are only apparent.

Understanding the work of Millikan and getting clear about its own internal difficulties and the limited relevance of its results for the authoritative prescriptivity of moral norms will be the central focus of this chapter. Only after making these things clear will Post’s own attempts to work central aspects of Millikan’s “teleosemantics” up into a

---

13 Post, *From Nature to Norm*, xvi.

14 Post, *From Nature to Norm*, 128.

15 Post, *From Nature to Norm*, ix.

metaphysics of morals be considered. In order to get clear on the terms of discussion and understand the problems Millikan has in view and how she attempts to surmount them, the first section of this chapter will canvass the central relevant features of Millikan’s thought. This is by no means an easy task. As one commentator astutely observes, Millikan’s central work ―is a hard book,‖ and Millikan herself complains that her work has ―managed to trip some very competent commentators.‖17 That said, the task is essential and not completely unmanageable—though we will have to hack through some pretty dense thickets.

After making the central strands of Millikan’s thought clear in Section I, in Section II we will consider some difficulties inhering in Millikan’s attempts to complete her own project in its own terms—a project in which she does not (so far as I know from my reading in the vast corpus of her work) directly consider questions of morality. Millikan’s project consists of a Herculean attempt to formulate a materialist theory of intentionality and representational content, and the “norms” she seeks to identify within evolutionary history are pressed into the service of her philosophy of language and mind in an attempt to fend off one of the most damning objections to materialist theories of representational content: How is it that a representation can be in error? Although the focus of this chapter is not to critique Millikan’s project, nor does it claim to be a definitive argument against Millikan’s project, weaknesses in that project are relevant in that they would undermine the groundwork of Post’s metaphysics of morals. As it turns out, Millikan’s project suffers from a number of serious flaws that raise questions about its adequacy as an answer to Millikan’s own questions. In what follows, these flaws will be grouped together under problems of the

17 Nicholas Shea, On Millikan (Belmont, CA: Thomson, 2005), 31; Millikan, White Queen Psychology, 123. Indeed, Millikan’s most frequent way of responding to pointed criticisms of her work is to complain that she has not been understood. Sometimes it seems to me that some of these critics have understood her work rather better than she allows.
that grounds proper functions, the problem of how it is that in humans the imperative has become separated from the indicative, and the temporal and logical difficulties in transmuting the diachronic causal skein summarized in language of “natural selection” into something that could rightly be called normativity.

The third section asks what limitations Millikan’s work would have as a grounding for moral norms if it succeeded (as I believe it does not) in grounding a materialist theory of intentionality and representational content. Here it will become painfully clear that Millikan would have no way to avoid moral reasons arbitrariness. On the one hand, the “biological norms” arise from causes that operate without regard to morality and cut across what many people would think would be the right thing to do. Papineau’s example of his knuckles being evolutionarily selected for hitting others—thus that being their “proper function”—is just one example out of a host that are ready to hand. But if we say (as we should) that such biological norms are not binding on us as people, then two consequences follow. First, by the same token, neither do “nicer” biological norms bind us as such. Neither some sort of “Golden Rule altruism,” nor putative features of a “moral brain,” nor some social instinct, has a more authoritative claim on us—qua “biological norm”—than the “biological norm” of hitting people with the knuckles that were designed to do just that. Second, if the biological norms do not bind us qua biological norms—if we, alone of all creatures we know of, are able to rise above them somehow—then what else will provide the needed authoritative moral norms? The norms manufactured by natural selection are (hopefully)

\[ \text{history} \]

\[ \text{separated} \]

\[ \text{transmuting} \]

\[ \text{natural selection} \]

\[ \text{normativity} \]

---

\[ \text{Note that throughout this chapter I am going to use this example of knuckles being “for” hitting rivals as a stand-in for whatever ugly elements of “biological design” arise from Darwin’s theory. Please note that I do not mean to be making a scientifically accurate claim that knuckles actually did evolve for this purpose (though perhaps they did, for all I know). If you don’t like this example, pick any unsavory element that evolution would have included as an element of the “biological design” of human animals. The point is what matters, not the specific example.} \]
irrelevant, and it doesn’t appear that we should look for help of some other form from that direction. In short, we seem either to be submerged in a nexus of biologically caused evolutionary “norms” that are frequently morally repugnant, or else to have risen above such biological facticity. The question then becomes: What else are we looking to besides nature to ground normativity? In that case, it turns out that the move from nature to norm needs a supplement from outside nature, materialistically conceived. This problem is only one of several that will be explored in depth in Section III.

In Section IV we turn at last to John Post’s attempt to craft a metaphysics of morals drawing on central elements of Millikan’s work. From the outset, having seen the difficulties Millikan’s work faces both in terms of her own project and when applied to morality, it is clear that Post faces an uphill battle. In addition to facing the difficulties stemming from his use of Millikan’s work, Post must also overcome a number of classic objections to any attempt to move from the facts of the natural world to the authoritative norms of morality. In the end, Post encounters serious difficulties in his efforts to craft a materialist metaphysics of morals that underwrites the authoritative prescriptivity of moral norms.

The final section returns to the problem of moral reasons arbitrariness developed in Chapter 2 and diagnoses the problems facing Post’s metaphysics of morals. We begin, however, with an attempt to understand Millikan’s difficult work in “teleosemantics” as the essential philosophical underpinnings of John Post’s attempt to move from nature to norm.
I. Millikan’s Teleosemantics and the “Normativity Problem”

“Reasoning, I insist, is done in the world, not in one’s head.” More than any other sentence I have come across in Millikan’s work, this statement captures the central thrust of her work. In this section I will try to show how Millikan attempts to make sense of a statement like this and why the idea here cuts to the very heart of her thought about representational content and intentionality.

One of the most difficult problems facing materialist attempts to account for representational content and intentionality is the problem of misrepresentation. The problem can be seen easily by considering causal theories of representational content. The central idea of such theories is that the representational content of our thoughts owes to the causes that give rise to them. For example, when someone thinks “snake,” the content of that thought is caused by the snakes that person has seen. The “snake” thought is about the snakes that have caused it. It seems pretty straightforward. But it turns out to be rather too straightforward, for all representational contents are thus caused and it turns out that no

---

19 Millikan, White Queen Psychology, 12; emphasis added.

20 Other candidate “if-you-can-understand-what-she’s-saying-here,-you’ve-really-got-it” sentences would be: “our meanings are as much theoretical items as are any other items,” (Ruth Garrett Millikan, Language, Thought, and Other Biological Categories (Cambridge, MA: MIT Press, 1984), 11); and “root purposing is unexpressed purposing; our job is to discover in what this purposing consists,” (Millikan, White Queen Psychology, 217).

21 The problem of “misrepresentation” and the very closely related “disjunction” problem are widely discussed in the literature, and are capably summarized in a number of places. One important statement of the problems is Jerry Fodor, “A Theory of Content, I: The Problem,” in A Theory of Content and Other Essays (Cambridge, MA: MIT Press, 1990). There is an outstanding extended treatment of this problem as it affects Millikan’s account of intentionality in Mark Perlman, “Pagan Teleology: Adaptational Role and the Philosophy of Mind,” in Functions: New Essays in the Philosophy of Psychology and Biology, ed. André Ariew, Robert Cummins, and Mark Perlman (Oxford: Oxford University Press, 2002).  Paul M. Pietroski, “Intentionality and Teleological Error,” Pacific Philosophical Quarterly 73, no. 3 (September 1992): 267-82, is a rigorous critique of Millikan’s account of representational content that takes the normative dimensions of the problem of misrepresentation as its starting point. Good summaries of the problem of misrepresentation can be found in Shea, On Millikan, 9-10; Millikan, White Queen, 7; Papineau, Philosophical Naturalism, 58; and Graham MacDonald and David Papineau, “Introduction: Prospects and Problems for Teleosemantics,” in Teleosemantics: New Philosophical Essays, ed. Graham MacDonald and David Papineau (Oxford: Clarendon Press, 2006), 4ff.
such content can ever be mistaken—for the content is nothing other than whatever happens to cause the “snake” thought. If a person thinks “snake” upon seeing a bit of rope coiled in the road at dusk (to borrow an example from Thomas Reid), then because the “snake” thought has been thus caused, that bit of coiled rope is properly part of the content of “snake” for that person. It is impossible to misrepresent something by the content of one’s thought, because those contents are about whatever causes them and all such content is caused by something or other. Whatever causes the content, that’s what the thought is about. Moreover, as Jerry Fodor points out, the content itself is intolerably disjunctive. On a causal theory, the content of a person’s “snake” thought is “snake – or – bit of coiled rope – or – dog’s tail – or…” The content of a person’s though is whatever does in fact (or, possibly, would in fact) cause that person to think “snake.”

The problem, as Millikan points out, not only affects causal theories of content, but also picture theories and rule theories (she calls them “PMese” theories due to the habit of her teacher, Wilfrid Sellars, of calling the symbolic logic of works like the *Principia mathematica “PMese”*). In each case, Millikan notes, the problem is that the theory, being materialist (at least in the forms she is willing to consider), does not have an account of normativity, and misrepresentation is obviously a normative notion. As she puts it, “each founders over the distinction between the facts of cognition and the norms of cognition”; Millikan aptly calls this the *normativity problem*.

---

22 Millikan, *White Queen Psychology*, 3-9. These various theories of content will not be a matter of concern here, but are brought up to introduce some central problems that Millikan is concerned to address.

Teleosemantics is built from the ground up to handle the normativity problem.\textsuperscript{24} The central idea is that normativity originates in the operations of natural selection in an organism’s evolutionary history.\textsuperscript{25} We have already seen the stock example used to introduce the central idea: the heart. Hearts have contributed to the survival of the organisms that have them by pumping blood. Given the evolutionary history of those organisms, that is what hearts are for. They have the \textit{proper function} they have because they have the history they have, namely a history in which a selective advantage accrued to the organisms that had them because they pumped blood.\textsuperscript{26} The heart may do other things, such as make a thumpity-thump sound or supply nutrients to a predator, but those are not \textit{proper} functions of the heart, since it was not its functioning in those ways that conferred a selective advantage in the evolutionary history of the organisms that have them. “Having a proper function depends upon the history of the device that has it, not upon its form of

\begin{itemize}
\item[\textsuperscript{24}]Indeed, although it is not clear that Millikan had this in mind from the beginning, she now insists that this is \textit{all} that teleosemantics does. Teleosemantics is a “piggyback theory” in which “You present your favorite theory of what a true representation is… [t]hen the teleosemanticist proceeds to explain what a false representation \textit{I} given your view. That is all teleosemantics amounts to,” Ruth Garrett Millikan, \textit{Varieties of Meaning} (Cambridge, MA: MIT Press, 2004), 66-7; Millikan’s emphasis.
\item[\textsuperscript{25}]Millikan is clear that she offers her definition of proper function not as an analysis of the concept of proper function, but as a theoretical definition that is justified by the weight it pulls in the theory of evolution. For reasons that will be made clear presently, Millikan has no patience with “conceptual analysis,” calling it “a confused program, a philosophical chimera, a squaring of the circle, the misconceived child of a mistaken view of the nature of language and thought [‘So, how do you really feel about that?’].” (Ruth Garrett Millikan, “In Defense of Proper Functions,” \textit{Philosophy of Science} 56 (1989): 288-302; reprinted in Colin Allen, Marc Bekoff, and George Lauder, eds., \textit{Nature’s Purposes: Analyses of Function and Design in Biology} (Cambridge, MA: MIT Press, 1998), 297. At the same time, Millikan hedges a bit in Millikan, \textit{Varieties of Meaning}, 36, note 4. Millikan’s way of proceeding is somewhat similar to what Richard Brandt called the method of “reforming definitions” in Brandt, \textit{A Theory of the Good and the Right}, 2-16. For a critique of the method of reforming definitions, see Nicholas Sturgeon, “Brandt’s Moral Empiricism,” \textit{The Philosophical Review} 91 (1982): 389-422. I have also criticized the method in Seeman, “Whose Rationality? Which Cognitive Psychotherapy?”
\item[\textsuperscript{26}]Millikan states this idea numerous places. The central place where she has developed this idea is her most important book, Millikan, \textit{Language, Thought, and Other Biological Categories}, hereinafter cited as LTOBC. For her fullest development of these ideas, see the “Introduction” and chapters 1 and 2 of LTOBC.
\end{itemize}
Thus, when a heart does not pump blood, it is failing to do what it ought to do, what it was biologically designed to do as that unfolded over its evolutionary history. An organism’s evolutionary history determines the normativity of its various biological functions.

Since the “normativity” in view here resides at the center of Millikan’s work and, indeed, is its whole point, her original (and, by her own estimation, still definitive) formulation of these ideas in *Language, Thought, and Other Biological Categories* merits some careful scrutiny. Here we find that “the business of the biological species of staying in business determines standards for individuals of that species, standards which, though they often correspond to averages, are not defined in terms of mere averages over the species.”

The standards for an organism and its various devices and behaviors are not a matter of what usually happens. To use two of Millikan’s favorite examples, the fact that most sperm

---

27 Millikan, *LTOBC*, 29. Millikan here (and in many other places) distinguishes her views on proper function in biology from those who think that such function should be identified on the basis of what the biological device currently does rather than on the history of what it has done. She thus stands within what has been broadly characterized as the *etiological* approach to biological function that traces back to the work of Larry Wright and is opposed by the *capacity* approach flowing from the work of Rob Cummins. A brief, helpful summary of these competing approaches may be found in William D. Casebeer, *Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition* (Cambridge, MA: MIT Press, 2003), 49-54. Although some have tried to combine elements of both camps, most theorists have fallen predominantly into one camp or the other. For a comprehensive overview of these issues and an introduction to the ongoing controversy, see Allen, *et al.*, eds., *Nature’s Purposes*. See also the collection by David J. Buller, ed., *Function, Selection, and Design* (Albany, NY: SUNY Press, 1999), especially Buller’s very helpful introduction, “Natural Teleology.” A fairly recent critique of Millikan’s etiological theory of proper function from a Cummins-style capacity perspective may be found in Paul Sheldon Davies, “The Nature of Natural Norms: Why Selected Functions are Systemic Capacity Functions,” *Noûs* 34:1 (2000): 85-107. Both, Perlman, “Pagan Teleology,” and D.M. Walsh, “Brentano’s Chestnuts,” in *Functions*, ed. Ariew et al. also bring in some elements from Cummins’ notion of “systems functions,” emphasizing more the standpoint of an engineer taking a synchronic look at a design problem than the diachronic, etiological approach favored by Millikan. In short, Millikan’s response to those who want to hold a capacity approach is that they have no plausible account of the normativity of proper functions, which is exactly what teleosemantists like Millikan hope a theory of proper functions can provide for a materialist theory of representational content and intentionality. Millikan’s argument against Cummins is laid out in Millikan, “In Defense of Proper Functions.”

cells do not successfully fertilize an egg does not mean that their proper function is not to fertilize an egg; nor does the fact that the camouflage of a caterpillar usually does not keep it from being eaten mean that the purpose of its coloring is not to conceal it from predators. The color of the caterpillar merely needs to have caused enough predators to have overlooked enough of its forebears that their coloring caused the coloring of the particular caterpillar in the present. The proper function of that coloring is to cause predators to overlook the caterpillar, even if, on average, it doesn’t usually work out for most caterpillars. Thus Millikan makes it clear that it is not averages or statistics that are in view in her account of intentionality, but “a specific sort of quasi-norm.”

It would be difficult to overestimate the importance of these “quasi-norms” for Millikan’s work. “The key notion that is needed in order to discuss intentionality (‘of-ness,’ ‘about-ness’) will be, in a way, only a by-product of the notion ‘proper function.’ This is a quasi-normative (roughly the biological or medical) notion ‘Normal’.” And so, when Millikan attempts to put “ten chapters in a nutshell,” we find the idea of “Normal explanation” featuring very prominently in the summary: “the sense of an indicative sentence is the mapping functions (informally, the ‘rules’) in accordance with which it would have to map onto the world in order to perform its proper function or functions in accordance with a Normal explanation.” As will be made clear shortly, the idea of sense—and thus the Normal—is central to Millikan’s theory of intentionality (she goes so far as to say that

---

29 Millikan, LTOBC, 5; see also p. 34.

30 Millikan, LTOBC, 5. The relation of “Normal” to the “proper” in proper function is not entirely clear in Millikan’s work; indeed, she sometimes seems to use them interchangeably.

31 Millikan, LTOBC, 11; emphasis in original. Although she later drops the practice, in LTOBC Millikan capitalizes Normal in order to remind readers that it is a technical, theoretical term that is not intended as a conceptual analysis and must not be understood in terms of a statistical average or preponderance.
“Fregean sense is the most basic stuff of meaning”\textsuperscript{32}, but first we must try to understand what Millikan means when she speaks of the Normal and the quasi-normativity embodied in that key term.

It is remarkable that, while the quasi-normative notion of the “Normal” is critical for Millikan’s account of intentionality, she never defines “Normal” in \textit{LTOBC} and in general does surprisingly little to make the idea clear. Particularly unclear is the relationship between the “Normal” and the “proper” in proper function. At times Millikan appears to use them interchangeably, as when she states that “Intentionality does have to do, very generally, with what is \textit{Normal or proper} rather than with what is merely actual. It also has to do with mapping relations—ones that are \textit{Normal or proper} rather than merely actual or average.”\textsuperscript{33} At one key juncture, however, Millikan uses “proper function” to define “Normal explanation”: “A \textit{Normal} explanation is a preponderant explanation for those historical cases where a \textit{proper} function was performed. Similarly, Normal conditions to which a Normal explanation makes reference are preponderant explanatory conditions under which that function has historically been performed.”\textsuperscript{34} This leaves the reader of \textit{LTOBC} in a bit of a quandary. Are “Normal” and the “proper” of proper function interchangeable? If so, how is it legitimate for proper functions to show up in a definition of Normal explanations? And why would Millikan call the Normal a “by-product of the

\textsuperscript{32} Millikan, \textit{LTOBC}, 6.

\textsuperscript{33} Millikan, \textit{LTOBC}, 86; emphasis added.

\textsuperscript{34} Millikan, \textit{LTOBC}, 34; emphasis added. See also the definition on p. 33.
But if they are not interchangeable, what exactly is the difference between them?

Initially it is tempting to read proper function as more basic than Normal explanation. The former is clearly meant as an ontological category, while it might seem that the latter is epistemic, being something that we do in noting and accounting for the design that does in fact exist in a biological device owing to its evolutionary history. This would ease—if not entirely resolve—the tension noted above between the “Normal” and the “proper” in proper function. However, in later work Millikan shuts down that reading of “Normal explanation,” making it clear that she did not mean “explanation” to be understood in an epistemic sense. In *Varieties of Meaning* she drops the “Normal explanation” terminology in favor of talking about “normal mechanisms,” having realized that her former terminology “caused some confusion, since many think of an explanation as being a set of propositions rather than what these propositions are about.” And so when Millikan spoke of “Normal explanations” in *LTOBC*, she apparently had nothing epistemic in mind, but rather the idea that the effects for which a trait has been selected are explained by the regular operation of a biological mechanism under Normal conditions. “Proper functions” are the effects for which a trait has been selected, and “normal mechanisms” (or “Normal explanations”) are the causal mechanisms by which the trait has caused those effects (and thus “explained” them) under Normal conditions. Thus Millikan talks of something being able to “fulfill through its normal mechanisms the function for which it

---

35 See the passage that goes with note 30 above.

36 Thus, “a function $F$ is a direct proper function of $x$ if $x$ exists having a character $C$ because by having $C$ it can perform $F$,” (Millikan, *LTOBC*, 26; first emphasis added.).

37 Millikan, *Varieties of Meaning*, 69, note 5. See also p. 85.
was selected.”\textsuperscript{38} It turns out that “proper functions” and “Normal mechanisms”\textsuperscript{39} are flip-sides of the same coin.

Applying this terminology to the camouflaged caterpillar, the theoretical account of the caterpillar’s color must refer to the diachronic causal skein of the phenotypic traits of the caterpillar’s forebear organisms, these traits being the “normal mechanism” (or “Normal explanation”) that \textit{caused} enough predators to overlook those forebears (in favor of the albino two branches over, say) so that our caterpillar has been \textit{caused} by the operation of those traits in the conditions its forebears typically found themselves in to have the camouflage colors it has. This \textit{effect} of those normal mechanisms—the camouflage coloring—is thus the proper function of the camouflage coloring\textsuperscript{40} that has been “selected for” over the course of the organism’s evolutionary history.\textsuperscript{41}

To this point, we have seen Millikan both (1) identify a serious challenge to materialist theories of representational content and intentionality, a challenge she calls the “normativity problem,” and (2) introduce the central elements of the “quasi-normativity” she will employ to address the normativity problem herself, namely, \textit{proper} functions and

\textsuperscript{38} Millikan, \textit{Varieties of Meaning}, 163. In this case the “something” to which Millikan is referring is what she calls a “pushmi-pullyu” representation, that is a representation where the indicative and imperative elements are joined together.

\textsuperscript{39} From here on I will drop talk of “Normal explanations” in favor of Millikan’s more recent choice of terminology.

\textsuperscript{40} This raises another issue that will be addressed below, namely the question of the degree to which a function should be taken to be proximate or distal. Speaking of the caterpillar’s “camouflage” understands the functions in a fairly distal manner, where there is a much more proximal story that could be told of the sensitivity of the typical predator’s sensory apparatus to color gradations and even the particular ways in which neuron’s in the typical predator’s brains are switched on and off in various patterns to trigger a mechanism of pursuit and consumption. Of course the causal story of the caterpillar’s coloring could also be told with more distal functions: the color caused the caterpillar’s forbears to survive or (at the limit) reproduce.

\textsuperscript{41} Millikan, \textit{Varieties of Meaning}, 68-9. Millikan follows Elliot Sober’s terminology in using “selected for” to refer a trait’s being selected because of some direct effect it has, rather than merely being correlated with something else that benefitted the organism.
normal mechanisms. In each of these notions it is the first part that introduces the quasi-normativity, both “proper” and “normal” referring to the causal factors in an organism’s evolutionary history that have caused, say, the caterpillar to have the coloration it has. Again, it is the proper function of the caterpillar’s coloration to camouflage it from predators because that trait in its forebears normally (again, not in the sense of statistically more often than not) caused predators to overlook them. Thus, due to that history, the coloration of the caterpillar is for camouflage, for making predators overlook it. From the diachronic causal skein of biology come quasi-normative purposes.

How does Millikan press this quasi-normativity into the service of a materialist theory of representational content and intentionality? It will be helpful here to return to Millikan’s attempt to put her central idea in a “nutshell”: “the sense of an indicative sentence is the mapping functions (informally, the ‘rules’) in accordance with which it would have to map onto the world in order to perform its proper function or functions in accordance with a Normal explanation.”42 Having considered the ideas of proper function and normal mechanism in this “nutshell,” we must turn to “sense” or “Fregean sense.” Not only does Millikan consider sense “the most basic stuff of meaning,” she even says that, “speaking loosely, sense is intentionality.”43 Clearly, “sense” is an important notion for Millikan.44

42 See note 31 above.

43 Millikan, LTOBC, 12; emphasis added. See also, note 32 above. Also, in a passage that identifies the (breathtaking) scope of her project and limns the basics of her approach, Millikan states:

I will place meaning and, in general, intentionality (aboutness, of-ness) in nature alongside sentences and the people who utter sentences. In so doing I will also try to show why sentences that exhibit subject-predicate structure, subject to negation, are of use to man, and how the law of noncontradiction (the essence of coherence) fits into nature. The notion Fregean sense is the basic tool that I will use in this construction (Millikan, LTOBC, 10; emphasis in original).

Obviously, “sense” is being asked to carry a lot of weight!
Millikan ties sense to “mapping functions” or “mapping rules.” These mapping functions “are rules in accordance with which a critical mass of sentences have mapped onto affairs in the world in the past, thus producing correlation patterns between certain kinds of configurations of sentence elements and certain kinds of configurations in the world, to which correlation patterns Normally functioning hearer interpreter devices are adapted, this adaptation explaining their successes.”

Now, there is a lot to unpack in that sentence, and to do so three more elements of Millikan’s system need to be introduced: real value (roughly, the “configurations in the world” mentioned above), transformations (roughly again, the “correlations patterns” above), and Millikan’s distinctive emphasis on the importance of consumer mechanisms (the “hearer interpreter devices”). To do this, it will be helpful to introduce Millikan’s trademark example: the dances of honeybees.

In order to indicate to other bees where honey is and direct them toward it, a honeybee performs an elaborate dance on a vertical wall of the hive, with orientation on the vertical axis indicating the direction of nectar relative to the hive and the position of the sun, and the number of “waggles” indicating the distance to the nectar (apparently the vigor of the dance indicates the amount of nectar available at the specified location, but that can be left to one side). One of Millikan’s most distinctive differences from others who had pursued materialist theories of representational content is her emphasis on the role of the

---

44 Millikan also states that sense, in turn, is “explicated by reference to the notion Normal,” (Millikan, LTOBC, 5). Ultimately, the quasi-normative is at the root of everything in Millikan’s thought, and the theoretical weight “sense” is asked to carry devolves onto Millikan’s quasi-normative notions: the Normal and proper function.

45 Millikan, LTOBC, 99.
consumer in creating content—in this case the “watching” bees.\textsuperscript{46} Previously theorists had focused on the need for the producer—the dancing bee, here—to represent things accurately. What Millikan argued was that the crucial element is the interpreter mechanism, for this is what will actually perform an action that can be selected for. The output from the consumer mechanism is crucial, as it is this that will either do or fail to do something useful that helps the organism survive and reproduce. “What we need to look at is the consumer part, at what it is to use a thing as a representation. Indeed, a good look at the consumer part of the system ought to be all that is needed to determine not only representational status but representational content.”\textsuperscript{47} Moreover, as Millikan makes clear, the producer and consumer mechanisms can either be two different organisms that have been designed to coordinate in their behavior (as with the bees), or these mechanisms can be within one and the same organism (as with eye-hand coordination in humans). In either case, the producer is actually subordinate to the consumer, the sole job of the producer being to produce representations of use to the consumer. “Although a representation always is something that is produced by a system whose proper function is to make that representation correspond by rule to the world, what the rule of correspondence is, what gives definition to this function, is determined entirely by the representation’s consumers.”\textsuperscript{48} The consumer will interact with the world and determine the functions that are selected for—or against.

\textsuperscript{46} The dance is actually performed in the dark, so presumably the dance is somehow felt rather than seen.


Thus, “the representation-producing side of the system had better pay undivided attention to the language of its consumer.” In short, the bee dance needs to send the other bees off in the direction of nectar.

The nectar in the bee-dance example is an element in what Millikan calls the “real value,” the second new notion needed to grasp her idea of “mapping functions.” “Real value” must be thought of in terms of the quasi-normative notion of a Normal condition. If there is no nectar and no sun out there in the world, then the proper function of the bee dance fails. The nectar, the hive, and the sun as the Normal condition of the bee dance form the representational content of the dance. It is of first importance to remember what “Normal” means here. In the evolutionary history of bees, the presence of nectar in the world and the ability to navigate toward it relative to the sun and the hive were conditions of the forebear bees’ traits causing behaviors that had a proper function—namely, bee dances bringing nectar to the hive. Or, try the same point another way around: think about what happens in the world that is advantageous to bees when they dance and respond to dances in the way they do: using the hive and the sun as fixed points of navigation, they get nectar! And that is indeed real value to them. More to the point, it was real value to the forebear bees, whose behavior under those Normal conditions of nectar being available and efficient navigation to it relative to the sun and the hive resulted in the survival both of them and their behavior, and this historical story makes the dances of current bees about those historically Normal conditions that made it advantageous to dance just so. Because of the

---


50 Actually, since bees are what Millikan calls pushmi-pullyu organisms, that is, organisms for whom representations are simultaneously indicative and imperative, it is not just the presence of nectar that is the real value of the bee dance (indicative), but also the flying of the consumer bees in response to the dance (imperative). For the purposes of this paragraph, the simplification will not be harmful.
evolutionary history of bees, the bee dance is *about* nectar, and about the hive and the sun by which consumer bees navigate toward the nectar; that is its representational content—a content that is *intentional*.

Finally, the bee-dance example helps make clear the importance of “transformations,” the third important notion in Millikan’s “mapping functions.” In the bee dances these transformations are a matter of how the relations between invariant aspects of the representation (the location of nectar relative to the sun and the hive) lead to or cause relations between aspects of the dance to co-vary (this particular bee doing a dance with 10 waggles 45 degrees off the vertical axis) such that there is a correlation or correspondence between the variant aspects of the dance and the real value of nectar, sun, and hive that the dance is about. Transformations in the invariant aspects of the Normal bee dance show up as transformations in the variant aspects of the bee dance, where these transformations are understood to be mathematical. “There are,” Millikan states, “operations upon or transformations (in the mathematical sense) of the icon [i.e., the representation] that correspond one-to-one to operations upon or transformations of the real value as such.”

There is thus a mathematically effected isomorphism between the waggles of the dance and its orientation relative to the vertical axis, on the one hand, and the location of nectar in relation to hive and the sun, on the other.

---

51 One of the few aspects of Millikan’s thought that I find hospitable is her adherence to a correspondence theory of truth and what she calls a “flatfooted realism.” See Ruth Garrett Millikan, “Metaphysical Anti-Realism?” *Mind* 95, no. 380 (October, 1986): 417-31; see also, Millikan, *LTOBC*, passim.

52 Regarding the invariant aspects of the representation, Millikan states that “there is no transformation of the bee dance that corresponds to replacing the sun with the moon such that the resulting dance maps the relation of moon, hive, and nectar,” (Millikan, *LTOBC*, 107-108).

With these ideas of “transformations” and “real value” in place, and with an understanding of why Millikan emphasizes consumer mechanisms over producers, it is now possible to make sense of Millikan’s notion of “mapping functions” and, in turn, the central idea of “sense.” When Millikan states that in a mapping function “a critical mass of sentences have mapped onto affairs of the world in the past,” she is tying the notion into the quasi-normativity she previously worked to establish. Just as the proper function of the caterpillar’s coloration is to camouflage it from predators because a critical mass of forebear organisms avoided being eaten in that way, so the bee dance (loosely, but analogously, a “sentence”) is about nectar in relation to the hive and the sun as its “real value” since a critical number of consumer bees were sent off in the direction of nectar because the dance produced by their conspecifics led to transformations in the consumer bees that were isomorphic to transformations in the world. (That sentence would do Millikan proud!) To put it more simply, a mapping function is that correlation between the behavior of the consumer bees and the world that is effected by the dance of the producer bee, a dance that is about nectar in relation to hive and sun because of past successful use of those conditions by a critical mass of “consumer” forebear organisms. These mapping functions are, in turn, at the heart of “sense.” In short, “the ‘sense’ of a language element is its Normal mapping rule.”

Millikan’s point is helpfully summarized by Nicholas Shea: “The content of a given representation is what the consumer system assumes it to mean. More precisely, a given representation is about the specific environmental conditions that were evolutionarily present [real value] when the consumer mechanism performed the evolved function of the

---

54 Millikan, LTOBC, 111.
activity triggered by that particular representation (i.e., the condition specific to the activity triggered by that representation)."\textsuperscript{55} It is the historically set Normalcy of the conditions—that transformations in the behavior of forebear bees (dancing \textit{so} in relation to the Normal conditions, \textit{including} the consumer bees) allowed them to \textit{use} relations in the world (the location of nectar in relation to hive and sun) in ways that were selected for—that sets what contemporary bee dance tokens are \textit{about}. The \textit{meaning} of the dance cannot be abstracted from the historically Normal conditions in relation to which bee dances can have a proper function. The meaning of a particular contemporary bee dance thus resides not in that dance itself (and certainly not in the bee’s head), residing instead in the Normal conditions of the bee dance, that is, in the bees’ evolutionary history. It is \textit{in relation to that history} that the bee could possibly make an \textit{error} in the dance or that consumer bees could make a \textit{mistake} in interpreting the dance. \textit{What the consumer bees would get wrong is not something in the producer bee’s head, but something in their evolutionary history}—the Normal conditions. And so we have an account of intentional representational content and of how those contents could \textit{misrepresent} what they are about. Bee dances are about their Normal conditions; secretions of adrenaline are about their Normal conditions\textsuperscript{56}; and—via a rather more complex linkage—thoughts about the transcendental deduction in the B-edition of the \textit{Critique of Pure Reason} are about their Normal conditions.

Recall how we opened this section: “\textit{Reasoning, I insist, is done in the world, not in one’s head.}”\textsuperscript{57} We can now begin to see what Millikan meant. Millikan is an inveterate externalist


\textsuperscript{56} Millikan, \textit{LTOBC}, 116-18, 239; Millikan, \textit{Varieties of Meaning}, 158.

\textsuperscript{57} See note 19 above.
and it is essential to her program (for obvious reasons) that she show that meanings “just ain’t in the head,” and in a much more fundamental way than Hillary Putnam was willing to entertain in his path-clearing essay.\(^\text{58}\) It should be clear that Millikan’s notion of sense that we have been laboriously unpacking completely displaces the classical thought that meaning is tied to intensions (with an “s”). Putnam, Donnellan, Kripke, and others had been chipping away at that notion for a few years, but Millikan thinks that they shrunk back from the implications of truly laying to rest the ancient ideal of a definition by which to set the extension of a term.\(^\text{59}\)

Millikan is determined not to shrink back, embracing with gusto all that she sees as “a necessary consequence” of her theory of intentionality.\(^\text{60}\) One consequence is the complete displacement of intensions as being the key to meaning, since meaning is finally tied to an organism’s history instead of its head. Millikan does not mince words here in staking out her opposition to what she calls “meaning rationalism.” “We cannot know a priori that we mean. Nor can we know a priori or with Cartesian certainty what it is that we are

---

\(^{58}\) Millikan skewers Putnam as a “meaning rationalist” (not, as we shall see, a compliment in Millikan’s world) in the “Epilogue” to \textit{LTOBC}. Putnam’s “Meaning and Reference,” as well as important essays by Kripke, Donnellan, and others, are collected in a superb volume with a very helpful introduction by Schartz, ed., \textit{Naming, Necessity, and Natural Kinds}.

\(^{59}\) One passage from the Euthyphro makes clear the idea of an intensional definition and all the very bad metaphysical stuff that it leads to—for which reason Millikan demands its complete and final extirpation.

Socrates: …for you agreed that all impious actions are impious and all pious actions pious through one form, or don’t you remember?”

Euthyphro: I do.

Socrates: Tell me then what this form itself is, so that I may look upon it, and using it as a model, say that any action of yours or another’s that is of that kind is pious, and if it is not that it is not.


\(^{60}\) Millikan, \textit{LTOBC}, 92.
thinking or talking about.... our meanings are as much theoretical items as are any other items.”\textsuperscript{61} To put this starkly, if you want to know what you mean, you have to investigate your evolutionary history rather than think about what you mean. “My hold on what I am thinking of depends upon how versatile and how reliable I am in performing acts of correct identification of my thought.”\textsuperscript{62} The idea that one can know what one thinks by simply reflecting on what is going on in one’s head has no place in Millikan’s theory of intentionality. Meaning arises in history and cannot simply be read off the detritus bobbing about the surface of our heads. As Millikan puts it, “root purposing is unexpressed purposing; our job is to discover in what this purposing consists.”\textsuperscript{63} “There is,” Millikan insists, “nothing diaphanous about consciousness.”\textsuperscript{64}

It is in this context that we meet “Swampman,” an infamous problem for theories like Millikan’s. Originally posed by Donald Davidson, the idea is that if an accidental double that was a molecule-for-molecule double of Davidson popped into existence, with all the biochemical switches in the brain in exactly the same position and so on, this double would not be thinking about anything or feeling any pain—even though the physically identical

\textsuperscript{61} Millikan, \textit{LTOBC}, 10-11; emphasis in original. Note that Millikan does say “that” and not “what” in this passage. In the omitted portion of this passage, Millikan also asserts that we cannot determine synonymy, ambiguity, or what is logically possible by reflection. Each of these is a matter for theoretical investigation of our evolutionary history. Similar passages in \textit{LTOBC} appear on pp. 92-3, 263-4.

\textsuperscript{62} Millikan, \textit{LTOBC}, 252. This investigation of the Normal conditions of one’s thought would have to include not simply an investigation of innate biological “hardwiring” but the “stabilizing proper functions” that elements of our language play in our culture. All of this must bottom out in a story of our evolutionary history, but Millikan is clear that the story is labyrinthine.

\textsuperscript{63} Millikan, \textit{White Queen Psychology}, 217. A few lines later Millikan declares that her “thesis will be that the unexpressed purposes that lie behind acts of explicit purposing are biological purposes.”

\textsuperscript{64} Millikan, \textit{LTOBC}, 92.
Davidson himself would be thinking of eating a Big Mac because he is hungry. The only difference between the double and Davidson is that Davidson has a history where the double does not; but the historical element of a teleosemantic theory like Dretske’s or Millikan’s implies—counterintuitively—that Swampman is neither thinking nor feeling hungry.

Millikan (characteristically, I am tempted to say) bites the bullet with relish. If you were to have a swamp-double, “that being would have no ideas, no beliefs, no intentions, no aspirations, and no hopes... This because the evolutionary history of the being would be wrong... That being would also have no liver, no heart, no eyes, no brain, etc. This, again, because the history of the being would be wrong.” Again, trying to make the point as clear as possible, the double would have no proper functions even though it functioned in every way just as you function. The processes in its brain would not be thoughts and they would not be about anything, since they would lack the evolutionary history that calls such intentional content into being. But Millikan has no choice but to chomp down on the bullet because her entire project rests on the quasi-norms that arise through an organism’s history of natural selection. And so Swampman has no thoughts about anything because the history of the world sets what thoughts are about.

---

65 Donald Davidson, “Knowing One’s Own Mind,” in Subjective, Intersubjective, Objective (Oxford: Oxford University Press, 2001); Dretske addresses this problem in Fred Dretske, “Representation, Teleosemantics, and the Problem of Self-Knowledge,” in MacDonald and Papineau, eds., Teleosemantics.

66 Millikan, LTOBC, 93; emphasis in original.

67 Proper functions are “defined in the end by reference to long-term and short-term evolutionary history, not present constitution or disposition. Were this not so, there could not be malformed hearts or nonfunctioning hearts nor could there be confused ideas or empty ideas or false beliefs, etc. Ideas, beliefs, and intentions are not such because of what they do or could do. They are such because of what they are, given the context of their history, supposed to do,” (Millikan, LTOBC, 93; emphasis in original).
“Reasoning, I insist, is done in the world, not in one’s head.” “The world”—an evolutionary history apart from one’s head—gives content to our thoughts, sets what they are about, and determines when they are mistaken. Meaning is a matter of theoretical inquiry.\(^{68}\) Just prior to this statement Millikan declares that her “desire is to kill meaning rationalism dead, and then beat on it.”\(^{69}\) The reason for her hostility should now be clear. If reasoning cannot be driven from the head and somehow placed in the world, it looks mysterious and not readily amenable to a materialist account of the mental. If, however, reasoning can be whittled down to mapping functions that have a quasi-normativity arising from their function in the diachronic causal skein that gave rise to an organism, the materialist can hold that intentionality—a seemingly intractable “mark of the mental”—marks nothing more than Normal conditions a biological organism’s forebear successfully used. “Intentionality is thus divorced from rationality, as sense is divorced from intension.”\(^{70}\)

Before moving on to a critique of Millikan’s account of intentionality and representational content, there are two last points of Millikan’s approach to intentionality that will become very important as we proceed. These need to be discussed briefly.

First, Millikan gives an immense taxonomy of various sorts of proper function in \textit{LTOBC}; an attempt to cover all of that taxonomy would lead us much deeper into some terminological thickets that are already quite dense. Fortunately, understanding the central distinction relevant to John Post’s attempt to derive a metaphysics of morals from Millikan’s

\(^{68}\) Millikan goes so far as speculate that “perhaps we needn’t take logic to be an a priori science at all. Insofar as logic deals with relations among concepts that have themselves been tested through experience, perhaps logic is an empirical science of sorts…. If logic cuts any ice at all, it \textit{is} ice about nature,” (Millikan, \textit{LTOBC}, 273; emphasis in original).

\(^{69}\) See Millikan, \textit{White Queen Psychology}, 12; emphasis in original.

\(^{70}\) Millikan, \textit{LTOBC}, 12.
work does not require going too much deeper into the bramble. The key notions are “relational proper function” and “derived proper function,” and we will see that Post leans heavily on them in *From Nature to Norm.*[^71] “A device has a *relational* proper function if it is its function to do or to produce something that bears a specific relation to something else.”[^72]

The common example is the chameleon, whose skin has a relational proper function of changing color to match the color of its background. A *derivative* proper function refers to the application of a relational proper function in a particular context. When the chameleon is sunning itself on a particular tree trunk, there is nothing in its evolutionary history that refers to that particular tree. Even the relational proper function is for matching *whatever* background at all the chameleon is on, and does not refer to this particular tree trunk. This is where derived proper functions enter into the picture. Deriving from the relational proper function of the chameleon’s background color matching mechanism as applied to this specific tree trunk, this particular shade of brown has a derived proper function of matching this particular tree trunk—and this despite the fact that nothing in the chameleon’s evolutionary history refers to that particular tree or perhaps even that particular shade of brown. Indeed, it could be that in all of that chameleon’s evolutionary history, that particular shade of brown has never been produced, and thus this particular function is completely new. This feature of derived proper functions features heavily Millikan’s attempts to account for human language and thought—in all of its vast differences from bee dances and secretions of adrenalin. Whether it can bear the weight Millikan puts on it is another matter. More central to the purposes of this dissertation, relational and derivative functions

[^71]: There is a brief, helpful summary of the ideas in Pietroski, “Intentionality and Teleological Error,” 270-71.

proper functions also figure heavily in Post’s account of moral normativity, serving as the biological provenance of the “bridge principles” that span the divide between nature and norms.\textsuperscript{73}

The final aspect of Millikan’s theory to be considered raises what I believe is perhaps the most difficult question facing her theory, and, really, any materialist account of things. Millikan notes that in lower organisms like bees or mice indicative representations and imperative representations are not separate, as they clearly are with human beings (at least to some extent). Millikan calls these “pushmi-pullyu” representations.\textsuperscript{74} Consider the bee dance again. The producer bee does not do one dance to indicate the location of the nectar and another dance to direct the consumer bees to “Go, now!” Nor to the consumer bees grasp the location of nectar and then decide what value to attach to that fact. They simply fly off in reaction to the dance in the indicated direction of the nectar. Or consider how the same bees will react to another “representation.” When a bee dies, it gives off a certain chemical that causes the other bees to drag it out of the hive. When a live bee is sprayed with that chemical, the other bees proceed to drag the live bee out of the hive. The chemical marker indicates “dead bee,” and this immediately directs “action”—even where the action is comical to us, who are able to understand the situation in a way wholly

\textsuperscript{73} In Post’s hands, Millikan’s key notions become “RFOR” (relationally for) and “CFOR” (circumstantially for). Post tells us that the key “bridge principle” MORAL (which itself twice uses the idea of something being “circumstantially for” doing something else) “is a special principle of bridge principle GOE,” which is about an adaptation being “circumstantially for” getting an organism to do something (Post, \textit{From Nature to Norm}, 131, 100). Any problems infecting Millikan’s account of “proper function” will devolve on the ideas at the very heart of Post’s work, unless he has ways of correcting those problems.

\textsuperscript{74} Millikan recognizes this connection in \textit{LTOBC} on p. 115, but she does not yet call them “pushmi-pullyu” representations.
inaccessible to the bee, certainly, and likely to any other creature at all.\textsuperscript{75} The bee has no awareness of its “goal state” in dragging a dead (or a live!) bee out of the hive, whereas we can see that dragging the live bee out of the hive makes no sense.

The question for the materialist, of course, is how we—alone, apparently, of all organisms—have come to be able to separate the indicative and the imperative. “How and why did perception-action cycles, which seem fully to characterize the cognitive character of the simplest animals, slowly give way to or become supplemented with more articulate and differentiated representations such as human beliefs, which are merely descriptive, and human desires, which are merely directive?”\textsuperscript{76} Here again, I submit, as with Korsgaard, we see an analytic philosopher bumping up against something very much like Heidegger’s Seinsfrage. Heidegger, of course, uses very different language, but the problem is similar: How do human beings transcend the immersion in Seiendes? How is it that in human beings the imperative has fallen away from the indicative, so that we are “held out into the nothing”?\textsuperscript{77}

These sorts of questions (though not in a Heideggerian argot) were raised recently by one commentator who managed to get under Millikan’s skin in a forum on Millikan’s Varieties of Meaning in Philosophy and Phenomenological Research. In his “Comments on Ruth Garrett Millikan’s Varieties of Meaning,” Jay Rosenberg notes that Millikan had promised in Chapter 16 of Varieties to address the transition from primitive pushmi-pullyu organisms to animals where the indicative and imperative have been separated out. “But Chapter 16, it turns out, has nothing at all to say about such a transition.” Instead Millikan merely assumes

\textsuperscript{75} It’s not clear that “action” is the right word here. That seems rather lofty for what is obviously the throwing of a chemical switch.

\textsuperscript{76} Millikan, Varieties of Meaning, 157.

the needed transition. The same holds true with respect to the provenance of detached representations of goal states (what the bees clearly were missing as they dragged their live comrade out of the hive). “That piece of the story,” Rosenberg points out, “remains permanently missing.”

In a rather terse reply to Rosenberg, Millikan admits that “Rosenberg is right that I don’t offer a theory of the origin of detached representations of goal states. I explain that they are not needed for the guidance of a great deal of surprisingly sophisticated purposeful activity, but not where they come from evolutionarily speaking.” Now one cannot demand everything from one author, and this—admittedly—is a pretty tall order (Heidegger spent his whole career on something very much like this one question). But, at the same time, this question is rather important for someone engaged in explaining language and thought as biological categories, and the fact that she doesn’t offer a theory of the origin of something one might rightly see as central to that which she is attempting to explain could understandably give a commentator pause.

The conclusion of Varieties of Meaning is fascinating to anyone who reads it with some Heideggerian sensibilities having been stirred by Millikan’s troubles over the separation of the imperative from the indicative. In the concluding pages of Varieties, Millikan observes that “possibly the most important achievement of theoretical thought,


79 Ruth Garrett Millikan, “Reply to Rosenberg,” Philosophy and Phenomenological Research 75, no. 3 (November, 2007): 701. While Millikan offers some intriguing examples from ethology, they are also clearly well short of anything like the sophistication one sees in human purposeful activity.

80 Without making a great deal of progress, as I see it. But he had the great merit of asking a really crucial question any materialist must answer and refusing to wave his hands over it.
resting directly on the capacity for language, is the capacity to represent historical time."\(^{81}\)

Heidegger’s early work focused on time as the key to the *ekstasis* of *Dasein*—the “standing out” or “standing apart” of human beings from the oppressively close causal skein of ontic beings that submerges all animals but ourselves. After the “turn” in his work, he focuses on language as the house of *Sein*, where *Sein* might be understood in part as that which is recognized when one sees that—despite the presence of the chemical marker—*this bee isn’t dead*. The later Heidegger, like Millikan, sees language as the key to this. The question, of course, is whether in doing so they have both run headlong into a “chicken-or-the-egg” problem: How did language happen without *Sein*—without an “aboutness” that goes much deeper than chemical markers and secretions of hormones?\(^{82}\)

Millikan concludes *Varieties of Meaning* thus: “The capacity to represent historical time gives rise to our ability to conceive of, to plan, and to carry out long-term projects that significantly change our environment. We quite purposefully and knowingly make what will exist in the future quite different from what has existed in the past.”\(^{83}\) That is an excellent statement of what wants explanation. Whether insisting that reasoning is *done in the world, not in one’s bead* gets much traction on this question may be a matter worth looking into a bit more.

---

\(^{81}\) Millikan, *Varieties of Meaning*, 226. On the next page Millikan states that “Grasp of historical time depends, in the first instance, on language.”

\(^{82}\) In my reading of an all too limited number of Heidegger’s works, the place where I think he makes his closest approach to this aporia is in Martin Heidegger, *Kant and the Problem of Metaphysics*, 5th rev. ed., trans. Richard Taft (Indianapolis: Indiana University Press, 1997), 140 where, in an extended discussion of something very like Millikan’s difficulty under the question of human “finitude,” Heidegger talks about time as “pure self-affectation.”

\(^{83}\) Millikan, *Varieties of Meaning*, 228.
Obviously there is more that could be said of Millikan’s theory of intentionality; it is the centerpiece of a subtle and immensely complex body of work. Millikan is a gifted and difficult philosopher. But enough has been done to make clear the general outlines of Millikan’s thought about intentionality, and to ask about some possible weaknesses in her account.

II. General Problems with Millikan’s System

I am not alone among Millikan’s readers in thinking that her theory of intentionality and representational content has done little more than change the subject. After Kenneth Taylor points out the vast differences between something like a rabbit thump signaling “danger” and human sentences that admit of many operations and transformations, Taylor draws his critique of Millikan’s Varieties of Meaning to a close by inviting “Millikan to clarify once more just what it is that bee signs have to teach about intensionality and what approaching rain clouds have to teach about the syntax, semantics and pragmatics of sentences like ‘It is about to rain in Palo Alto.’ I myself can think of very little.”84 Indeed, one can read Millikan’s work and just be left entirely cold by the whole project, with all the machinations and subtleties looking like so many diversions from the main problems, which simply have not been addressed.85 But not everyone will have this reaction, and it will be helpful to attempt to draw out the specific problems with Millikan’s project itself before moving on to consider the difficulties her project would create for moral philosophy if one

---


were to accept it. If there are difficulties that Millikan’s project faces on its own terms prior to its application to moral philosophy, those problems would carry over into a project attempting to ground moral philosophy on the affected aspects of Millikan’s work.

A. Jackdaws, Bees, and Hormones; or, The “Quasi-” in Millikan’s Quasi-normativity: Why We Should Keep It and the Problems It Creates for Millikan on Intentionality

We have seen that in the first few pages of Language, Thought, and Other Biological Categories Millikan speaks of the biological world as affording “quasi-norms”: “The key notion needed to discuss intentionality… is a quasi-normative (roughly the biological or medical) notion ‘Normal’. 86 Unfortunately, the “quasi-“ more or less disappears from Millikan’s talk about norms and bio-norms and the like after the initial pages of LTOBC. To be sure, she will sometimes recall to the readers’ minds the limitations of the “normativity” underlying all her work, as when she states her preference for talk of “semantic mapping functions” 87 rather than “semantic rules” because “the notion of a ‘rule’ tends to have prescriptive overtones that I wish, by all means, to avoid.” 88 Or, again, “the norms for language are uses that have had ‘survival value’, as Sellars put it. As such these norms are indeed disposition-transcendent, but they are not ‘fraught with ought’. They are not prescriptive or evaluative norms.” 89 Most of the time, however, Millikan simply talks about

86 Millikan, LTOBC, 5; see note 30 above.

87 See pp. 395ff. above on the centrality of “semantic mapping functions” for Millikan’s attempt to account for intentionality.

88 Millikan, Varieties of Meaning, 76, note 6.

norms, purposes, and what is normal with no “quasi-”, no scare quotes, and no “schnormativity.” As a consequence it becomes very easy to forget the limitations of her project—even if it were entirely successful on its own terms.

The exceeding thinness of Millikan’s account of the normativity needed to get her story of representational content and intentionality off the ground is clearly seen when one stops and considers her examples. My personal favorite is Millikan’s story of some jackdaws that had imprinted on the ethologist, Konrad Lorenz. Apparently, the jackdaws would attempt to drop worms into his ear, owing to the shape, size, and color of the opening being such as to trigger the mechanism in them to feed their chicks.90 Clearly there is a purely causal mechanism at work here, there being no reason for the jackdaws to act in this way. The behavior is caused by mechanisms at work in them. Of course, this satisfies Millikan’s own account of meaning, intentionality, and misrepresentation, since there is an evolutionarily established “normal” mapping function that set the “meaning” of the jackdaw behavior. It is “about” feeding chicks, and in the case of Lorenz’s ear, they “misrepresent” the ear as a baby jackdaw gullet. Remember, also, the bees dragging their living compatriot out of the hive because it has been sprayed with a chemical that triggers “drag-it-out-of-the-hive” behavior. Indeed, even an adrenaline rush is “intentional” in Millikan’s parlance, this secretion having the appropriate evolutionary history to qualify.91

---

90 Millikan, *Varieties of Meaning*, 171.

91 Millikan, *Varieties of Meaning*, 156; Millikan, *LTOBC*, 117. In *LTOBC* she calls an adrenaline rush an “intentional icon”; in her later parlance she calls an adrenaline rush a “representation.”
But of how much interest is any of this to an account of intentionality—the mental representation of either a present or non-present or non-actual object as that which some thought or act is about? For starters, there is clearly nothing intensional (with-an-‘s’) on the part of the jackdaw in its behavior; the intensionality, insofar as it is there at all, is supplied by the “normal conditions,” and the jackdaw either is or is not (as in the case of Lorenz’s ear) in relevantly similar conditions. 92 If the jackdaw is in conditions relevantly similar to “normal conditions,” then it drops worms into its chicks’ gullets, having “represented” things correctly. If the jackdaw is not in relevantly similar conditions, then it “misrepresents” the situation and drops worms into Lorenz’s ear. But the “intensionality” resides in the conditions, not in the jackdaw. The jackdaw itself has no criterion by which it distinguishes what it should or should not do in a situation. We humans, on the other hand, do. That is why we find something comical about the image of a jackdaw dropping worms into Lorenz’s ear, or a bunch of bees dragging a “dead” bee out of the hive kicking and screaming. We supply an intensionality (with-an-‘s’), a criterion for meaning in the case of the behavior, that the jackdaw has no awareness of or possession of in any way. For the jackdaw the “criterion” resides not in itself or in anything it is aware of, but in a set of facts that unfolded within its evolutionary history and the conditions in which it now finds itself caused to behave in relation to the “normal conditions” of its forebears. The causal mechanisms in a particular jackdaw will interact with the conditions into which it has been

92 Millikan’s emphasis on the importance of these conditions is helpfully summarized in Pietroski, “Intentionality and Teleological Error,” 271-2; and Shea, On Millikan, 28-34: “Her [Millikan’s] idea is that what a representation stand for cannot be found in the function of any system, but instead in the conditions in the world that must be in place if the system is to function as designed,” (28).
thrown in a set way outside the jackdaw’s awareness or control.\textsuperscript{93} Such intensionality (with-an-‘s’) is very thin indeed. And, \textit{pace} Millikan, it is difficult to see in any of this a bridge or precursor to human intensional understanding of criteria for our actions. In that sense of “aboutness,” jackdaws, bees, adrenaline secretions, and squirrels are non-starters. The location of the putative “intensions” is all wrong, residing not in the jackdaws, but in the conditions into which they are thrown. Those conditions impinge \textit{causally} on the jackdaw behavior. Mapping functions are considerably different than meaning, since the latter involves \textit{an organism} (humans are the only ones known to be capable of doing it) applying a criterion by which a thought or behavior is rightly applied in a situation as that which should be thought or done. In Millikan’s account, the “normativity” is in the wrong place—and she offers no solution for how evolutionary processes might move it from the conditions to the organism itself. Because of this, Millikan’s charming examples of animal behavior do not serve as precursors or bridges (even conceptually, much less historically) to the human phenomena she set out to explain.\textsuperscript{94}

But things are worse still for the “normativity” in Millikan’s account of intentionality (with-a-‘t’ now). The behavior of jackdaws, bees, and the non-human biological world in general displays little if any resemblance to human intentionality at the crucial points. Simply put, Millikan’s account never transforms the mechanistic causality that is all Darwin’s story affords us into the “shoulds” and “oughts” of human normative experience, and it fails to account for the diversity of objects to which that normativity applies. Starting with

\textsuperscript{93} In the language of Heidegger and Sartre, it is thrown into a facticity from which it cannot emerge. There is neither responsibility nor awareness; nothing ontological, only the ontic.

\textsuperscript{94} Millikan’s solution, as we have seen, is to assert that meaning and intentionality are “in the world” and not “in the head.” Whether this squares with the phenomena themselves is another matter.
the second point first, humans think normatively about many objects that are not only not present (whether if it’s raining in Palo Alto I should travel there), but also about non-actual objects (how many wings I should draw Pegasus as having; how I should sum 37 and 52; what should I do if a trolley were…).\textsuperscript{95} Compare this to the jackdaws. The jackdaws interact only with the present facts of the situation; if the jackdaws do have some sort of tenuous “normativity” to their behavior, it is only about the facts of the situation in the present. They interact causally only with a present opening of roughly such-and-such size and shape and coloring. They do not represent a possible opening in the future or get nostalgic for the times gone by when their chicks would open their beaks just so. They merely are caused to do something in response to a certain present stimulus. A jackdaw doesn’t think about what it ought to do if some human being’s ear were presented to it. Nor could it think about what it should do if some snake were to evolve that laid its eggs in jackdaw nests, swallowed the worms offered to its appropriately shaped mouth after hatching, and eventually ate all the jackdaw hatchlings before eating “mama” and “papa” too and departing the nest. We human beings, however, can imagine ourselves into Mama Jackdaw’s situation, assess the (non-actual) situation where there is such a snake, and decide what should be done in the situation. This human capacity to abstract from the present and actual is radically discontinuous from anything we find in the non-human biological world. Indeed, even the most advanced animals do nothing like what human beings do routinely: “the use of signing in apes is restricted to situations in which the eliciting stimulus and the reward are clearly specified and present, or at least very close…. the ‘meaning’ of an ASL

\textsuperscript{95} Perhaps the last two examples should be grouped as “non-present” rather than “non-actual.” In any case, however one classifies things, the point is still clear. The human repertoire of normative thinking has a variety of objects that extends well beyond anything encompassed by Millikan’s examples.
sign to an ape is simply the episodic representation of the events in which it has been rewarded.\textsuperscript{96} Apes, along with the rest of the non-human world, are submerged in the facticity of the situation into which they have been thrown, interacting \textit{causally} with \textit{actual} features of the conditions in which they find themselves, rather than \textit{normatively} with what they take to be the reasons that either are or would be found for a thought or action in a present, non-present, or even \textit{non-actual} situation. Human normativity is applied across a range of objects that far outstrips the very narrow range of object with which animals interact causally.

Yet as wide as the gulf is between the objects with which animals interact causally and the objects humans interact with normatively, this is not the widest of the chasms separating animal “normativity” and human normativity. Wider still is the divide between the mechanistic causality operative in Millikan’s “normativity” and the “should’s” and “ought’s” of genuine human normativity. Where what needed to be explained was humans having \textit{reasons} for their thoughts and actions, we are offered an account of animals driven by mechanisms that \textit{cause} their behavior. While an explanation of the latter is fascinating and important, it cannot stand in for an explanation of the former. As seen above, at the heart of Millikan’s attempt to transmute evolutionary causes into normative reasons for what “ought” to be believed or done, is her account of “normativity” as arising from an evolutionary history that turns mere functions into \textit{proper} functions. But, as Jerry Fodor rightly notes, “in evolutionary theory as elsewhere, if you wish to deploy the idiom of posed problems and designed solutions, you must say something about what designing requires \textit{over and above mere causing}. Lacking this distinction, everything a process causes is (vacuously)

one of its designed effects, and every one of its effects is (vacuously) the solution to the problem of causing one of those." In Millikan’s account it is the evolutionary history of a function that transforms it into a proper function, thus licensing her to talk in terms of the normativity of design (what the heart should do) rather than in mere causal terms. The crucial question is whether evolutionary history can bear the weight Millikan needs it to bear, effecting the needed change from causality to normativity. If not, we still have a merely causal account after all, and Millikan fails to provide the normativity essential to any account of representational content and intentionality.

There is, of course, a crucial constraint on a materialist account of intentionality: the account cannot itself lean on the intentionality of which it is supposed to be providing an account. Fodor again gets this right: “if your goal is a reductive theory of intentionality, then your account of this difference [between designing and merely causing] cannot itself invoke the intentional idiom in any essential way." Any reifying of “Mother Nature” or “Natural Selection” simply will not do. Millikan does sometimes succumb to such reifying, and it makes her account very slippery. Thus one finds her talking in some rather startling ways: “what interested natural selection in selecting…,” “cognitive systems are designed by evolution to…,” “would seem much the most likely to be preferred by natural selection…,” and (my favorite), “That sounds like a lot of concepts and abilities to be developed all at once. But Nature has apparently handed us quite proficient inborn abilities tailored to…”

97 Fodor, “Deconstructing Dennett’s Darwin,” 177; emphasis in original.


99 Millikan, Varieties of Meaning, 82; Millikan, White Queen Psychology, 11; Millikan, Varieties of Meaning, 84; Millikan, LTOBC, 312; complete with the capital “N” in the original.
Of course Nature, natural selection, evolution, and the like did nothing of the sort, and the intentional idiom applied in this way can mislead both author and reader.

However, problematic as Millikan’s incautious application of intentional language to Nature is, this is not the chief route by which Millikan smuggles normativity into her account so as to change causality into normativity. The main point of entry can be seen in the passage near the opening of Language, Thought, and Other Biological Categories where she lays out her “key notion” as “a quasi-normative (roughly the biological or medical) notion ‘Normal’.” It’s the “or medical” that is of particular interest here, for the medical is, of course, quite different than the biological understood in terms of the forces operating in the diachronic causal skein that gives rise, in time, to all biological diversity. Indeed, the notion of the medical makes a rather prominent appearance at the beginning of Aristotle’s Nicomachean Ethics as an example of a specifically human art or practice of teleological interest alongside the practice of shipbuilding and the art of waging war. And it is the telos or aim in medicine as a human practice that makes medicine of such interest to Aristotle.

But Darwin will countenance none of this; his is a mechanistic account where efficient causes reign unchallenged, final causes having been routed from the field. Evolution is driven by a diachronic causal skein. And it is part and parcel of the materialism to which Millikan is committed that it follow in this vein, explaining everything in terms of purely natural stuff that gives rise to all that is through efficient causality. By installing the medical at the heart of the “key notion” that vouchsafes the normativity from which everything else in her teleosemantic account of intentionality flows, Millikan relies on a teleologically-laden notion that is glaringly out of place in her putatively materialist account.

100 See note 30 above; emphasis added.
One can begin to wonder if it is a human stance—Darwin’s mechanistic biological world viewed through the lens of a particular interest we bring to it—that is doing the heavy lifting in Millikan’s account of intentionality. If so, then the intentionality of the natural world derives from human intentionality after all, and the latter has not been explained by Millikan in materialist terms.

Recall the problem facing Millikan: How can normativity arise from (efficient) causality? Her solution hinges upon the evolutionary history of an organism giving rise to (quasi-)normativity by transforming functions into proper functions. But what does this history consist of? It is tempting to say “one damn thing after another,” but the “damn” has to drop out, since it is an evaluative element. The evolutionary history with which Millikan must work is nothing but a diachronic causal skein free of all evaluative elements. Here the point Fodor raised is crucial: What story, over and above mere causing, does Millikan have to tell?

Let’s return to Millikan’s examples and see what we find. The heart is for pumping blood because it was by performing that function that hearts contributed to the fitness of the forebears of the organism that has the heart. Of course this is a purely causal process, one that we conveniently summarize by saying that the heart is for pumping blood. What has actually gone on is that a particular organism had a heart that, say, pumped blood in slightly higher quantities than other organisms did, thus delivering oxygen to muscles a little more rapidly and allowing the organism to escape a few more predators or beat a few more competitors for females into submission. The details of the story don’t matter much here (and I’m clearly making no claims for the historical accuracy of the story I have just told). The important thing is that a phenotypic variation caused a particular organism to do
something that caused it to reproduce at a slightly higher rate than other, competing organisms. That causal story repeated countless times has caused your heart to pump blood.

At no point do we need to bring in the normative to give the account in question. The causal story is sufficient and completely accurate; there is a complete mechanistic story of the causes of the heart that never appeals to anything normative or even quasi-normative. The normative story is a kind of shorthand for the true story—the causal story—that is what actually happened. In short, the normativity is adventitious, being brought to the causal skein by human beings in making sense of what has gone on in the causal realm. There is nothing normative that actually went on in the evolution of hearts or of anything else; the Darwininan story is a mechanistic, causal story. Period. That we find normative stories of the heart being for pumping blood to be helpful is an interesting fact about us, not about what actually went on. It is interesting that we find a fairly undigested normative story very helpful in thinking about the biological world, and that such stories open up further vistas of understanding in similar terms—but Darwin’s point was that such teleological

101 Dennett makes a point much like this in Daniel C. Dennett, The Intentional Stance (Cambridge, MA: MIT Press, 1987) and Dennett, Darwin’s Dangerous Idea. Of course the issue here is whether the causal story can explain the stance itself, the intentionality of which is not the less real for the fact that the materialist has to bring it to the causal story. One cannot explain the normativity of the natural world through an intentional stance and then turn around and use the natural world to explain the normativity essential to the intentionality of the stance. Fodor makes a not unrelated point (but from the side of the natural processes rather than the ―stance,‖ as it were): “Dennett often writes as though ‘Mother Nature’ is just a figurative name for natural selection. That is, of course, perfectly okay; but you can’t both talk that way and also hold, as a substantive thesis, that her intentionality explains its,” (Fodor, “Deconstructing Dennett’s Darwin,” 187; emphasis in original).

102 This, of course, was a point Immanuel Kant made in Kant, Critique of Judgment. The teleological stance is a fact about us. The fact that it is we who bring that stance to the mechanistic world means that the teleology we find there cannot be used to explain the origins of the teleology in our intentionality, but is, rather, heautonomous.
stories in the biological realm could be retold in purely mechanistic terms. In short, the origin even of the quasi-normativity in Millikan’s account is us.

But say I’m wrong about this, and that there is some sort of “normativity” in the biological world owing just to the history of this or that particular organism or mechanism (and not our way of viewing that history). Still, let’s be very clear about the severe limitations on any such “normativity” at the heart of Millikan’s attempt to do justice to intentionality and to representational content that can genuinely be mistaken. First, there is nothing in the causal biological world Millikan accepts that genuinely responds to a should or an ought. It cannot be emphasized strongly enough that (in Millikan’s mechanistic world\(^{106}\)) the heart does not pump blood because it should, but because it was caused to do so. It is blind—wholly unresponsive—to any such norms, responding only to actually present causal forces.\(^{105}\) Indeed everything in that mechanistic, causal world—with the glaring exception of us\(^{106}\)—operates without any regard to the oughts and shoulds of the normative world in which we

\(^{103}\) Quine was correct in stating that the point of Darwin’s theory was to drive teleology out of the biological world, bringing it under a mechanistic view of the world.

\[^{104}\] Efficient cause… gained the upper hand over final cause with the rise of physics…. Final cause still had its explanatory duties too, not only in relation to the mind of man but also in biology, where it became an embarrassment, depriving biology of the austerely scientific status that physics had come to enjoy. Darwin at length settled that matter, reducing final cause in biology to efficient cause through his theory of nature selection (W.V. Quine, The Pursuit of Truth, Rev. ed. (Cambridge, MA: Harvard University Press, 1992), 75; quoted in Post, From Nature to Norm, 90).


\(^{106}\) I do not, of course, mean by this to be saying anything about quantum indeterminacies at the micro-level. Clearly, the materialist will have a story to tell at that level as well that keeps anything “spooky” at bay.
live and move and have our being. Millikan’s “normativity” is a normativity where nothing responds to the “norms” supposedly shot through the biological world. Again, it looks a lot like a “normativity” where nothing is responding to the “norms” as norms, the normativity being simply a heuristic overlay brought to the causal forces that are really operative. In any case, a “normativity” where nothing is aware of or responding to the “norms” is an extremely limited “normativity” at best.

Second, and a point that is prominent throughout the critical literature on teleosemantics, the “normativity” found by Millikan and other teleosemanticists is blunt in the extreme, especially as compared with the extraordinarily fine-grained nature of the representational content for which it is supposed to give an account. A detailed investigation of the literature and the attempts to respond to it is not in order here, but it will be helpful to give a brief summary of the problem facing teleosemantics and anyone who proposes to use central aspects of it to ground a moral theory. The ubiquitous example in the literature is taken from an experiment that showed that frogs would snap their tongues not only at bugs flying in the vicinity, but also at bee-bees researchers tossed in the air near the frogs. Cutting right to the point of all the ink that has been splattered on this example, the problem for Millikan and friends is that where representational content requires a very specific attribution of the content (“bug”) the evolutionary functions afforded by an etiological account allow for a variety of different possible specifications of

\[107\] In one way or another, the problem of the mismatch between the bluntness of the representations sanctioned on a teleosemantic account and the fine-grainedness of representational content comes up in each of the following, thought-provoking sources: Fodor, “Deconstructing Dennett’s Darwin,” 176-84; Walsh, “Brentano’s Chestnuts,” 317-24; Pietroski, “Intentionality and Teleological Error”; Perlman, “Pagan Teleology”; Bernard Enç, “The Indeterminacy of Function Attributions,” in Functions, ed. Ariew et al.; and Taylor, “Sign, Sign, Everywhere a Sign!” The problem here is very closely tied to the problem of misrepresentation.
the intentional content (“bug,” “bug-or-bee-bee,” “amino acid packet,” “affordance of survival benefit,” etc.). D.M. Walsh summarizes the problem this way: “If the intentional content of a state is determined by its evolutionary function, then evolutionary functions must be at least as finely demarcated as intentional states.” The problem is that there is a “disparity between the looseness of evolutionary functions and the stringent requirements on any account of intentional functions.” There is an obvious disparity between the blunt “normativity” derived from the causal history of evolutionary functions and the fine-grained normativity of representational content. The basic idea of the problem is pretty simple: You have one evolutionary function that could be mapped onto quite a number of different possible intentional contents. Again, the function of snapping at a fly can be cashed out intentionally as snapping a fly or a small, dark moving thing or a protein packet or… and on it goes. But the evolutionary function can’t distinguish between these—it’s compatible with any of them. What, then, is making the distinction? It can’t be the evolutionary function. It’s ham-fisted. Where the intentional contents can be very nuanced, the evolutionary functions are blunt and indistinct. It’s not clear how the big clump of the one could underlie all the fine distinctions of the other.

Moreover, the ways of trying to match up the evolutionary functions with supple mental phenomena can look implausible and quite ad hoc. There is no obviously principled way to specify where to land on the “proximal”/”distal” scale of the attribution of content;

---

108 Walsh, “Brentano’s Chestnuts,” 317, 318; emphasis in original. Walsh is very helpful on this point, and I recommend his essay. Walsh’s own solution is just to swallow down intentionality whole as “a sui generis realm of wholly natural, irreducibly teleological categories… Aboutness is real, but it is not really something else,” (335). This reminds one of a comment Fodor makes: “If we did come across a miracle, we’d call it a basic law or an anomaly and swallow it down accordingly. (Cf. electricity, action at a distance, and what goes on inside black holes),” (“Deconstructing Dennett’s Darwin,” 173). Fodor’s own way of summarizing the same problem Walsh has in view is rather more technical: “selected for’ is likely to be transparent to the substitution of necessarily coextensive terms, whereas ‘realizes,’ ‘thinks about,’ ‘believes that,’ and the like certainly are not,” (“Deconstructing Dennett’s Darwin,” 183).
that is to say, there seems to be no reason to prefer the most “distal” attribution of content ("affordance of survivability benefit") or the most “proximal” ("neural event ‘n’"). Both extremes are unacceptable as the frog’s representational content, with the proximal side of the scale undermining the very possibility of misrepresentation. Perlman notes that the “problem is, if you are willing to go a little proximal, why not all the way? You cannot have your only reason for stopping in the middle be that this still retains the possibility of malfunction, and thus misrepresentation.” Once again, it can look quite a bit like it is the theorist who is bringing the normativity to the diachronic causal skein she is considering and finding that there is no unique way to “gild and stain” (as Hume would have it) the causes with the normativity.

If that is right, then clearly the teleosemantic theorist has made no progress explaining the normativity by appealing to the causality operative in the selective history of a heart or whatever it is she is attempting to explain. But even if that is wrong, it is still the

109 Casebeer, *Natural Ethical Facts*, 49-52, has a helpful overview of the “proximal”/“distal” distinction in which he relates the proximal end of the continuum to Robert Cummins’ more synchronic “systems” or “capacity” approach to functions, and the “distal” pole to Larry Wright’s more diachronic “etiological” approach to functions. We have obviously been looking at the latter approach in some depth, since Millikan is a committed adherent of this view. The Cummins-style approach is more like what an engineer might do by just analyzing a system in the present to see how it works. This approach has a very difficult time giving any account of error or misrepresentation, since there is no standard of what it is to go right or wrong (except the one in the engineer’s head, and clearly that won’t do for a materialist account). The function just is whatever it is doing. Of course this is where etiological accounts appeal to selective history to establish a standard—what is “normal.”

110 Perlman, “Pagan Teleology,” 280-89.

111 Perlman, “Pagan Teleology,” 283.

112 The problem in this paragraph is related to questions about specifying the domain of what Millikan calls “local natural information” in *Varieties of Meaning*. This adds some twists that would be interesting to explore, but there is no need to do that here. For an overview of some of these issues, see Taylor, "Sign, Sign, Everywhere a Sign!" and Millikan's reply. Also informative is Nicholas Shea, “Consumers Need Information: Supplementing Teleosemantics with an Input Condition,” *Philosophy and Phenomenological Research* 75, no. 2 (September 2007): 404-35, and Ruth Garrett Millikan, “An Input Condition for Teleosemantics? Reply to Shea (and Godfrey-Smith),” *Philosophy and Phenomenological Research* 75, no. 2 (September 2007): 436-454.
case that the bluntness of the “normativity” that Millikan attempts to identify within the selective history of organisms is what we would expect from the diachronic causal skein that is an organism’s selective history. It is far from clear how such blunt “normativity” could explain the fine-grained normative distinction involved in representational content capable of being wrong about bugs and bee-bees.

A third difference between Millikan’s “normativity” and the genuine article is that there is no prescriptivity, no address that directs that something should or ought to be done. Recall Millikan’s own insistence that the “norms” she finds in the biological world “are not prescriptive or evaluative norms,” and that she abjures any “prescriptive overtones” which are something “that [she] wish[es], by all means, to avoid.” Millikan’s “normativity” does not direct anybody to do anything, but rather is meant to serve the very limited purpose of underwriting the possibility of misrepresentation that is so obviously a feature of mental representation. But there is a question of whether a “normativity” with no prescriptive element is worthy of the name, even in scare quotes or with a “quasi-“ attached. Once again, the fact that Millikan finally has only causality to work with in her story of an organism’s selective history creates the problem. It would seem that misrepresentation requires some notion like prescriptivity; something about the perceptual situation rightly should have been perceived otherwise than it was. A prescriptive element of the situation uniquely directs the one who is representing the situation toward this content rather than that content. Justification is an issue. Now, again, clearly we have questions of justification in view in the situation somehow. But Millikan is attempting to show how the causal story of

113 See notes 88 and 89 above.

114 This issue is not unrelated to questions of motivational internalism, a position many philosophers have found attractive.
selective history works something very much like prescriptive normativity into the mental, and clearly her account cannot appeal to that which she is proposing to explain.

Millikan must therefore derive the needed normativity from the history of the causes that led to the survival of the organism. Starting with hearts again, the causes of the heart’s pumping blood do not address hearts with reasons to pump blood; they simply cause it to pump blood as it does. At any point in the causal chain we find only causes, and not anything like prescription. The same holds for the present heart; it was caused to pump blood thus and so; and if it does not pump blood it is not failing to do anything it should do, except from our standpoint as we bring a medical stance to it. We are the ones looking to the diachronic aspect of the causal story out of our interest in telling a normative story about what hearts should do, but there is no prescriptivity in the causal events themselves. Hearts—and hormones and bees and frogs and everything else in the evolutionary stories available to Millikan—are not addressed by a selective history, but only by the present causal forces as they have been shaped by myriad other causal forces in the past that determine them to be one way rather than another. We take an interest in the history; the present heart itself is responsive only to causal forces in the present. Nothing addresses shoulds or oughts to anything, and (as we saw above in the first difference between Millikan’s “normativity” and actual normativity) nothing is responsive to should or oughts.

With all these differences between Millikan’s “normativity” and the real thing—differences clearly seen in the liberality with which she is able to apply it to things like adrenaline rushes, bee dances, pumping hearts, splashes of beaver tails, and the like—it is hardly surprising that some have thought that she simply changes the subject. Such “normativity” is indeed quasi-normativity. One might even think that to grace it with the
word normativity at all, even with scare quotes, quasi’s, or as *sch*normativity, is ill-advised. Perhaps it is so different as to be unworthy of the name at all. Moreover, it may well be that whatever traction Millikan does get in solving the normativity problem, she gets by smuggling in the normativity by tacit appeal to a stance (“biological or medical”) that takes an interest in a history of the causes of biological mechanisms being what they are and then brushes normativity onto that history. If something like this is going on in Millikan’s account of representational content and intentionality, then obviously there is a vicious circularity afflicting the account. We are left with the need to explain the normativity that human beings bring to the diachronic causal skein. How did *we* acquire this normative capacity?

Faced with such problems, Millikan can look suspiciously like an eliminativist with respect to that which does not fit into her theory. Under the cover of her attack on “meaning rationalism,” it can seem that Millikan is *eliminating* normativity, meaning, and—therefore—intentionality: “we are no more in a position to know merely via Cartesian reflection that we are truly *thinking*, i.e., that we or our thoughts intend anything, than that we are thinking truly. Absolutely *nothing* is guaranteed directly from within an act of consciousness.”

This is one way to solve the problem of misrepresentation and the attending problem of normativity that gives the former problem its teeth. Perhaps Millikan is not an eliminativist with regard to intentionality, meaning, and normativity, but, in any case, it is clear that the “normativity” she finds in the biological world is very thin indeed. Indeed, insofar as it resides in the diachronic causal skein of an organism’s evolutionary history, it really looks nothing like normativity at all. It is brushed with normative hues by

---

115 Millikan, *LTOBC*, 92; emphasis in original. It’s not clear that the word “Cartesian” is doing anything here other than serving as an evocative trope: “It’s *Cartesian*, so it must be bad.”
human beings who bring a normative heuristic to the diachronic causal skein. If that is the
case, Millikan has explained intentionality not as a biological category, but only by appealing
to human intentionality itself.

B. Problems with Anchoring “Normativity” in History

This subsection will briefly draw out some additional problems for Millikan’s
etiological account of functions, problems that face someone who wishes to ground a
“metaphysics of morals” in Millikan’s work.

We have already encountered “Swampman” above, the objection being that an exact
physical duplicate of Davidson that popped into existence as a massive quantum freak with
no evolutionary history would not have thoughts or experiences. As Fred Dretske notes,
“To describe this result as a problem for teleosemantics may be too generous. Some would
describe it as a reductio ad absurdum.”116 Perhaps. But I would like to point out that the
problem behind Swampman is not merely a conceptual problem that can be swept aside by
saying that it is the product of philosophers’ nasty habit of doing conceptual analysis and
coming up with conceptual problems.117 Swampman serves to clarify a problem, but the
problem is a theoretical problem (as well as a conceptual problem) that has its roots not in

116 Dretske, “Representation, Teleosemantics, and the Problem of Self-Knowledge,” 75. Dretske
goes on to lead the reader through an elaborate series of what Dennett would call “intuition pumps” to show
why the Swampman consequence of teleosemantics isn’t so bad after all. It is not necessary to plunge into all
the literature surrounding Swampman to draw attention to the problems of interest here.

117 See Colin Allen, Marc Bekoff, and George Lauder, “Introduction,” in Nature’s Purposes: Analyses of
Function and Design in Biology, ed. Allen et al., 13; and Millikan, “In Defense of Proper Functions.”
fanciful “barn façade” kinds of counterexampling, but in an element of Millikan’s own theory.\footnote{118 I am not meaning to knock barn façade counterexamples in the literature on epistemological externalism. I rather like them. It’s safe to say, however, that Millikan does not.}

In etiological accounts of proper function, history makes proper function. The Swampman objection trades on this fact. But there are all sorts of (less spectacular) examples of the same problem in the biological world. Any phenotypic expression of a genotypic variance (the word “mutation” seems out of favor in the literature) suffers the same problem. A structurally different fin may allow a fish to swim faster while expending less energy, but—according to Millikan’s etiological account of proper function—the fin’s new structure has no proper function. Call this the prototype problem.\footnote{119 Although he heads in a different direction than the argument here, there are similarities between the prototype problem and a problem raised in Paul Sheldon Davies, “‘Troubles for Direct Proper Functions,’” \textit{Naïs} 28, no. 3 (1994): 370ff.} Indeed, if that fish breeds before ever having escaped a predator because of the speed difference (it would have escaped with the old fin), the same fin structure of the mutated fish’s spawn will not have any proper function either. But as soon as one fish (perhaps several generations hence) with the different fin structure escapes because of the speed advantage conferred by the fin, then the progeny of that fish will have fins with a proper function. A question arises whether this would also confer a proper function on the same fin structure of “aunt” and “uncle” and “cousin” fishes that have not had ancestors that escaped because of the fin structure.\footnote{120 And what of the forebear fishes? Would their fins retroactively gain a proper function?} They would lack the needed history, so it would seem that their fins would have no proper function, while the structurally identical fins on their “cousins” would have a proper function. Additionally, however many times the prototype fish escapes because it can swim...
faster, its fin will never have a proper function because its fin lacks the requisite history. Only the fins of its progeny will have a proper function of swimming faster. Tying the advent of proper functions to history in this way leads to results that seem arbitrary and counterintuitive.

But perhaps the prototype problem need not bother us all that much, for eventually it will all come out in the wash—all the fins having a proper function of swimming faster over the course of enough generations. The initial arbitrariness eventually works itself out. But the problem intensifies as we move to the relational and derivative proper functions that become important for Millikan’s account of language and Post’s account of morality. Causal factors spread out across two generations determine proper functions, and they can do so in wildly divergent ways that shift from one generation to the next. Say a business owner in an underdeveloped nation survives to reproduce by being honest and thus gaining a good reputation in the community. That behavior would set a derived proper function for his children. Say that his son adheres to his proper function of being an honest shopkeeper and starves, but his daughter is conniving and underhanded and survives to reproduce. Now her children will have conniving and underhanded behavior as a proper function. Proper functions will change rapidly. Moreover, any behavior that contributes to a selective advantage will have a proper function for the next generation. Say that someone has a pattern of obsequious behavior that allows him to survive as a slave and reproduce

\[\text{121} \] It’s harder to pin down exactly the role of history in Millikan’s understanding of relational and derivative proper functions. Another way to think of this is to return to the chameleon example of relational proper function. The chameleon has a relational proper function of turning the color of whatever ambient background predominates; thus a particular chameleon has a derived proper function of turning this particular shade of brown. And so human beings might have as a relational proper function (with a nod to Gibbard) a propensity to use language to enlist the service of other human beings in their projects. A slave and a master, in their very different “Normal” conditions, would have very different derived proper functions—much like two chameleons, one sitting on a brown trunk and the other on a green leaf.
successfully where more restive slaves are killed. Then obsequious behavior has a proper function so long as the conditions are “Normal.” There may also be coordinated proper functions along the lines of Hegel’s or Nietzsche’s master/slave relationship.

The point is that because Millikan ties proper function to evolutionary—and, derivatively, cultural—history, proper functions can vary wildly from generation to generation, and from person to person and can have any content, so long as it contributes historically to fitness. Proper functions vary along with causes, taking whatever shape they are caused to have. And these causes are not responsive to reasons.

C. How Is It that in Humans, Alone of All Animals, Imperatives Have Come Uncoupled from Indicatives?122

Although Millikan acknowledges “fundamental” differences between the “representations” animals have and human beliefs, it is clear that she means to shed some light on the latter. So, on the one hand, in “Biosemantics” one finds Millikan specifying “six very fundamental ways” in which human representational content is unlike that of animals, and in Varieties of Meaning Millikan acknowledges that there are “ways in which the representational capacities of humans may differ from those of nonhuman animals, certainly in degree and probably in kind.”123 On the other hand, the title of Language, Thought, and Other Biological Categories promises big things, and chapters are dedicated to crafting biologically plausible accounts of negation, the act of identifying, and the law of non-contradiction. Indeed, in the introduction to LTOBC, we find Millikan laying out her ambitions:

122 See note 106 above.

123 Millkan, “Biosemantics,” 97; and Millikan, Varieties of Meaning, 212.
I will place meaning and, in general, intentionality (aboutness, of-ness) in nature alongside sentences and the people who utter sentences. In so doing I will also try to show why sentences that exhibit subject-predicate structure, subject to negation, are of use to man, and how the law of noncontradiction (the essence of coherence) fits into nature.  

That is aiming pretty high. And Millikan’s ambitions have not diminished in her later work:

The central question that I want to address is how and why, during the evolution of perception and cognition, organisms have acquired inner representations that are more sophisticated than pushmi-pullyu signs [that is, signs where imperatives are set by indicatives]. How and why did perception-action cycles, which seem fully to characterize the cognitive character of the simplest animals, slowly give way to or become supplemented with more articulate and differentiated representations such as human beliefs, which are merely descriptive, and human desires, which are merely directive?  

Such is the program Millikan sets for herself, and there is no doubt she is playing for all the marbles: She is pushing as hard as she can to give a materialist account of human representational content and intentionality, or at least to open a path toward it.

We have already seen that, under pressure from Rosenberg, Millikan backs off her claims a bit: “Rosenberg is right that I don’t offer a theory of the origin of detached representations of goal states.” And it will be helpful to make plain just how far she is from realizing anything like these ambitions.

Perhaps more than any other philosopher, Martin Heidegger made clear just how vexing the problem of “detachment” of imperatives from the facticity of indicatives really is

---

124 Millikan, LTOBC, 10. See note 43 above.

125 Millikan, Varieties of Meaning, 157; emphasis added. Note that the difference here seems not to be a kind difference, but a matter of degree. Rosenberg’s critique of Millikan draws attention to this.

126 See notes 78 and 79 above. Millikan continues on to insist that her speculations about the animal world show that such detached representations “are not needed for the guidance of a great deal of surprisingly sophisticated purposeful activity,” (Millikan, “Reply to Rosenberg,” 701; Millikan’s emphasis). Millikan clearly is trying to chip away at the uniqueness of what goes on with human intentionality and representational content. It’s unfortunate that Millikan chooses to be dismissive of much of Rosenberg’s critique (her admission in the face of Rosenberg’s pointed critique is the most substantive moment of the whole book symposium).
for materialists (though he would not have used this language). This “detachment” in all its manifestations is related to what Heidegger puzzled over for his entire adult life as the question of Sein.\(^{127}\) The question is one of “clearing” or ekstasis—standing back from the facticity that is oppressively close to all the animals of which we are aware. Human beings are the only Dasein of the biological world, having somehow made beings present there with themselves as beings “lit up” in a world. Squirrels do not decide about (imperative) a tree as a tree (indicative), but are merely swept along in a flood of pushmi-pullyu representations. They are submerged in the flood of Seinde im Ganzen, the ontic serial of things—the causal skein we have seen Millikan trying to overcome by an appeal to history. Human beings—somehow—have become detached as the causal skein has “fallen away.”

Like Millikan, Heidegger also tried time first as an answer to the Seinsfrage. For the early Heidegger, time as ek-stasis is crucial for Sein, being that by which Dasein stands out from the skein of beings to have a place, an indexical “here,” from which to think and decide about them. With respect to both the past and the future that Dasein has been enabled to bring into the present, Dasein finds place from which to be “here” with the present entities, rather than being submerged without remainder in the crushing flow of them in an oppressively close skein of things following one another. The present only becomes such as the past and the future are brought into it.

But to grant the self time is just to beg the question. It’s like just assuming a stance with respect to the accreted causes that meet in a particular evolved organism and thereby thinking that progress has been made in showing how that stance itself arose from the causal

---

\(^{127}\) I do not mean to say that Millikan and Heidegger were pursuing identical problems.
skein. Precisely in examining the historical and making it relevant to the beings in question, the most basic question has been begged.

The later Heidegger, after the so-called *Kehre*, tried to approach the *Seinsfrage* by talking about language as “the house of Being” and the advent of Being in the fundamental *poesis* of language which first makes trees (those particular nodes of causality that our ape ancestors climbed up, slept in, and swung from but were never aware of as trees) by naming them as something that is not self. *Aboutness* was thereby first opened up. Before science, before logic, must come

the breakaway of humanity into Being. In this breakaway, language, the happening in which Being becomes word, was poetry. Language is the primal poetry in which a people poetizes Being. In turn, the great poetry by which a people steps into history begins the formation of its language. The Greeks created and experienced this poetry through Homer. Language was revealed to their Dasein as a breakaway into Being, as the formation that opens beings up.¹²⁸

Poetry is transcendence of the serial of entities, the initial freeing or “breakaway” from animal submersion into Being that first opens a being to be *here* as standing out from and amidst beings that are not “I.”

But this too is just to grant oneself the whole problem—the basic question is simply begged. Let me allow Jerry Fodor to put the point in his own inimitable way. In a critique of Dennet’s tendentious use of Darwinism, Fodor hammers away at the mismatch between the bluntness of the putative representational content of evolutionary functions and the extraordinary fine-grainedness of mental representations, Fodor observes in a footnote that Dennett blithely dismisses questions of whether the frog has an intentional object when it snaps its tongue: in effect, Dennett says, “Who cares whether it’s a bug or bee-bee or

affordance of survival benefit that the frog snaps at? There just may be no fact of the matter.” But, Fodor notes, Dennett is then left with the problem of why “the same wouldn’t hold of intentional ascriptions to *us*.129 Then we come to the point: “Do not reply ‘that’s because we have language’ or I shall break down and commence to gnaw the rug: Exactly the same indeterminacy would infect the assignment of meanings to reliably coextensive predicates of, say, English.”130 The problem is with these “intentional objects” and the fine-grainedness of our representational content of them: “This oak tree here that I am climbing.” How did I come to stand over against this oak tree here instead of that maple that I was hiding behind yesterday? As Fodor points out memorably enough, you can’t just say “It’s because of language.” This obviously begs the question.131

Now, it may seem that we have wandered quite a distance from Millikan, but the concluding words of Millikan’s “Précis of *Varieties of Meaning*” should dispel that quickly:132

If language, like light,133 is a medium of direct perception of the world, then the identities of objects and properties that have already been discovered by one’s language community are manifested in a simplified way through language. The additional perspective on the world that linguistic communication makes possible also allows development of the representation of linear historical time. Unlike the animals, we understand that in acting we are constructing new parts of history, not merely circling through the same times again. This allows us purposefully and

---


131 As I said above, my own take is that Heidegger gets closest to the full gravity of the problem when he speaks of “pure self-affectation” in Heidegger, *Kant and the Problem of Metaphysics*, 140. There is a helpful exposition of this idea in Richardson, *Heidegger: Through Phenomenology to Thought*, 110-117. Finally, I think, all the talk of self-affectation is just to say that it just happened and we don’t really know how. Fodor again: “If we did come across a miracle, we’d call it a basic law or an anomaly and swallow it down accordingly. (Cf. electricity, action at a distance, and what goes on inside black holes),” (“Deconstructing Dennett’s Darwin,” 173). See note 108 above. Of course, it is no part of this dissertation to argue this point.

132 One could also consider the passages quoted at notes 81 and 83 above.

133 This simile closely parallels the way Heidegger talks about language and *Sein*. 
knowingly to make what will exist in the future quite different from what has existed in the past. Animals don’t understand the forward dimension of time.\textsuperscript{134}

This is a clear statement of the problem, but, as Rosenberg points out in his admirable critique of Varieties of Meaning, this question is not so much addressed as begged by Millikan, and most of Varieties is spent emphasizing the likenesses with the animal world and obscuring the differences. Millikan’s “theoretical perspective of biological functionalism… teaches us many likenesses. But it is evidently also a perspective which tempts one to the ‘irrational exuberance’ of treating such likenesses as identities, and that is a temptation to which Millikan, I think, too often succumbs. My complaint is that many equally instructive and important differences thereby go missing.”\textsuperscript{135}

The biggest difference is the truly astounding fact that we think about objects (even non-existent ones) that we represent as standing over against ourselves, swimming, in media res, in language and time. The Seinsfrage—the question of detachment—is the question that must be answered by the materialist who would account for language and thought as biological categories, and this is the question that, glaringly, is unanswered, hardly even approached, by Millikan. Her copious and fascinating stories from the animal world serve mainly to distract attention from the real question that needs to be answered: Just how did the “pushmi” and the “pullyu” come apart in us, alone of all the biological world? Here is a huge difference between all the stories about animals and what needs to be explained about human representational content, fundamentally marked by time and by the normativity and distance

\textsuperscript{134} Ruth Garret Millikan, “Précis of Varieties of Meaning,” Philosophy and Phenomenological Research 75, no. 3 (November 2007): 662; emphasis in original. There are too many connections with Heidegger in this passage for me to explore, so I’m just going to hope that enough has been said to make some of them plain at this point.

\textsuperscript{135} Rosenberg, “Comments on Ruth Garrett Millikan’s Varieties of Meaning,” 700; emphasis in original.
of language capable of being *about* beings as beings. In obscuring the differences, it can seem rather like Millikan is changing the subject.

### III. Problems with Applying Millikan’s System to Morality

The point of the previous section was to highlight some of the difficulties besetting Millikan’s project in its own terms. Perhaps Millikan or someone else might be able to offer convincing responses to these problems, but there seems to be a tremendous amount of work to be done. My own assessment is that Millikan has not set off down a promising path in solving the problem of intentionality that she wants to address.

In this section the point is to consider difficulties that might attend an attempt to apply Millikan’s thought to moral philosophy. So far as I know, Millikan herself makes no such attempt. Thus, the present section is not a critique of Millikan’s thought as such, but rather a consideration of the difficulties that might be involved in extending Millikan’s thought into an area of philosophy where she has not worked. Nor does the present section critique the specifics of John Post’s attempt to formulate a “metaphysics of morals” from the mechanisms giving rise to the “normativity” Millikan devises in addressing her own questions. Rather, this section attempts to point out some more general reasons why Millikan’s work might strike one as an improbable place to look for a way to ground morality.

#### A. It’s a Long Way from Millikan’s “Normativity” to Moral Norms

Millikan herself emphasizes that the “normativity” she attempts to locate within the causal world of evolutionary history is not prescriptive, and I have tried to make plain just
how wide that divide is. The first thing to point out in drawing attention to the difficulties in attempting to move from Millikan to morality is that where it is clear that Millikan’s “normativity” is not prescriptive, it is equally clear that moral normativity is prescriptive. The experience of moral norms involves being addressed as a person with reasons that direct one as to what either should or should not be done and rightly demand that one comply. This recognition of independent, authoritative prescriptivity addressed to us in reasons is of the essence of moral norms (as was argued in Chapter 2). But this is just what we don’t find in Millikan (as she herself insists, and Post also recognizes). Indeed, what we find instead is that she encounters serious difficulties in her attempts to overcome the causality of the biological world to vouchsafe even the highly attenuated “quasi-normativity” she hopes will make some progress in explaining the possibility of misrepresentation in a materialist theory of representational content. The robust, genuine normativity involving prescriptivity is not even in view in Millikan’s materialist project (even though it should be, if my arguments in the previous section are correct). And this robust, prescriptive normativity is precisely what characterizes moral norms.

136 There are, of course, much bigger issues lurking in the area here that I cannot address without doubling the size of this dissertation. G.E.M. Anscombe puts her finger on a key question that would need to be asked if the critique developed in this dissertation were to open out into a positive thesis that might inform a moral philosophy. Speaking what could be meant by a moral “ought,” Anscombe notes that the idea cannot be separated from a “law conception of ethics”: “For its suggestion is one of a verdict on my action… And where one does not think there is a judge or a law, the notion of a verdict may retain its psychological effect but not its meaning,” (Anscombe, “Modern Moral Philosophy,” 252). A law requires a law-giver and an authoritative prescription requires a prescriber with rightful authority. Pursuing this clue had been part of the original scope of this dissertation, but, well, one cannot do everything in one dissertation—and this one is ambitious and lengthy enough as it stands.
One key aspect of moral normativity is that it addresses people with reasons, and Millikan’s “bionorms” do not do this.\footnote{This element of prescriptivity is what Gibbard tries to capture with his “norm-expressivism,” but, as argued in Chapter 3, he encounters severe difficulties in trying to capture any sort of substantive objectivity for his views. One would seem hard-pressed to avoid moral reasons arbitrariness by moving in that direction.} Instead, causal forces in the evolutionary history of organisms set the various proper functions of organisms, never addressing them as persons, but causally impinging on organisms on a biological level. This is true even of relational proper functions (remember the chameleon example). Relational proper functions owe to the causal history that gave rise to them. Thus the variable pigmentation of the chameleon’s forebear organisms obviously caused a sufficient number of predators to pass over them and prey on those organisms that lacked the phenotypic difference. And the derivative proper functions (matching that shade of brown making that bird miss it) are derivatively caused in the interface of that organism with its conditions.

These same causal forces will be at work in the relational and derivative proper functions that govern human behavior within culture. With a nod to Gibbard, the relational proper function (if it is such) of using language to enlist the cooperation of others whose assistance I need in securing the goals of my own projects has been caused by a massive history of death, survival, and reproduction. Now when the human organism is thrown into a culture of highly educated American social democrats in the year 2009, that organism will have a derivative proper function of going along with that culture and with other groups he meets along the way (the partying friend of the girl to whom he is attracted). But the same relational proper function of a human organism thrown into Germany in 1933 will cause a different derivative function. The same might be said of a (possible) relational proper function (again with a nod to Gibbard) of capitulating to conversational pressure; the
derivative proper functions will be caused by the varying cultural situations in which human organisms find themselves. Organisms are impinged upon causally rather than addressed prescriptively with normative reasons.

Additionally, one can also start to see that when Millikan’s “normativity” is brought over into the moral realm, it gives rise to a very unattractive sort of moral relativity in which moral “reasons” are causally beholden to amoral forces and vary accordingly. Human organisms have been caused to have certain dispositions to react in particular ways in particular situations, some of these dispositions being proper functions because of the role they played in forebear organisms. These proper functions will vary with the evolutionary history of particular organisms—even more given the variability of derivative proper functions that arise when the relational proper functions of human organisms are thrown into the welter of conditions characteristic of human culture. And, so long as the relational proper function is operating in “Normal” conditions (which can be as morally ugly as you like, since “Normal” is just whatever situation where this particular adaptation caused the organism to survive\textsuperscript{138}), the derivative functioning will have all the “normativity” Millikan’s materialist etiological account of proper functions can provide.

But such relativity does not characterize moral normativity, which at least in its intentions reaches out beyond dispositions and causes to what should be done and what is in fact wrong if it is done. Indeed, we regularly take moral norms to fly in the face of all sorts of

\textsuperscript{138} Thus, there is no reason why “Normal” conditions for the human organism’s treatment of step children might not be situations of scarcity where they are competing for resources with one’s own biological progeny. Or, again, it is perfectly possible that “Normal” conditions are those of slavery, or of living in conditions of competition with various “out-groups,” or all sorts of possible causal situation which causally filtered what would and would not survive. Again, whatever passed through these filters, whatever traits caused the forebear organisms to survive and reproduce down to the present cause of this particular organism, just that is the proper function of all mechanisms in that organism. Millikan’s “normativity” can take us no farther.
diverse moral behavior, sweeping it aside as simply irrelevant to the issue of what is right and wrong. *Whatever* the causal forces that lead someone abduct a young girl, pump her full of drugs and beat her with a hose in the basement of a brothel until she “services” thirty men a day, no matter how such behavior may be properly functioning when considered as a derived proper function in these particular conditions understood as “Normal” given the history of human evolution, the acts are just wrong.\textsuperscript{139} Such is the force of a moral norm—and people do not in fact take that force to vary with causes or to be required to wait on judgment until all the noses are counted.

That brings us to a last, and very important, difference between Millikan’s "normativity" and genuine, prescriptive moral normativity. Moral norms are taken to have authority, whereas there is no plausible ground of authority for Millikan’s “norms.” Once again, the key thing is that Millikan has nothing other than pure, efficient causality to work with, and her “normativity” makes only the most tenuous of progress—if any at all—beyond that causality. What is important is that causality lacks authority, and that it seems that we are able to see through any of Millikan’s quasi-norms and cast them off as irrelevant.\textsuperscript{140}

Perhaps the clearest way to see this is thinking of how you might react to someone who is arguing that what should have been done in a situation is something that you think is wrong. What would you do if she advanced a host of *causes* that explained what was in fact done? What would you do if she advanced *reasons* that justified what was done? In the former case,

\textsuperscript{139} I do not, of course, mean to say that Millikan would say otherwise. I do mean to say that if we carry her notions of biological “normativity” over into the moral realm, those biological “norms” on their own do nothing at all to underwrite what we do in fact know of moral norms from our experience of them.

\textsuperscript{140} If so, this would apply not only to bad “norms” like hitting someone with knuckles in an attempt to gain access to a female, but also to good “norms” like some sort of tit-for-tat “altruism.” This is tied to the issue of the “detachment” of the “pushmi’s” from the “pullyu’s” that will be considered below.
one dismisses the causal explanation as missing the point of the situation, which required a reasons justification. In the latter case, one argues against (or sees the correctness of) the reasons, trying to make plain why they are wrong, and why a different position is justified. Causes lack authority; they possess only force. And whatever is caused is not thereby justified. If it is the case that Millikan’s “norms” do not effectively transcend the causal account of selective history as I have argued, then they lack the authority essential to moral reasons.

B. “Pushmi’s,” “Pullyu’s,” and Morality

For all organisms, either imperatives and indicatives come apart, or they do not. It is clear that for us, to some significant degree, they come apart. We have already seen that Millikan runs into severe difficulties in advancing her own project of giving a materialist account of representational content at this juncture. It also turns out that an attempt to press Millikan’s thought into service of moral philosophy would encounter problems at this point.

On the one hand, to the degree that the “pushmi’s” and the “pullyu’s” remained joined within Darwin’s story of natural history, the moral philosopher has to take the bad with the good. If there are perhaps some inherently social drives that push us toward some kinds of cooperation, there are perhaps also drives to discriminate against members of an “out-group” or to abuse step-children and the like. The “purposes” or “proper” functions or “norms” arising from the biological world as Darwin painted it can be expected to reflect a

---

141 So far as it truly makes sense to employ language of indicatives and imperatives to organisms like bacteria.
“Nature red in tooth and claw,” as Tennyson put it. To the degree that the imperatives have not become detached from the indicatives of that factual world, the “bionorms” will cut across what many have considered to be the best of human moral thought and action at some bizarre (and quite possibly hideous) angles.

Millikan’s “pushmi-pullyu” organisms are submerged in Darwin’s biological world, being swept along by their evolved drives and unable to stand back from the forces operative in the situation to identify objects about which they can reason and decide what ought to be done with respect to them given the past and looking forward to the future.

The biological world that sweeps these “pushmi-pullyu” organisms along is not a pretty world, but is a long, grinding biological “arms race.” In that arms race, any function that causes an organism to survive and reproduce will be a proper function. If that is a “spider hawk’s” larva eating the non-vital organisms of the paralyzed spider first, then the larva is functioning properly when it does that—and, since it is a “pushmi-pullyu” organism, it will do exactly that unless there is some sort of mechanical malfunction. If it is hitting a rival with knuckles to gain access to a (or yet another) female, then the “pushmi-pullyu” organism will commence with hitting the rival in “Normal” conditions. That just is the world in which Millikan’s “normativity” operates; and for “pushmi-pullyu” organisms, this dictates what will be done (barring mechanical malfunction). Given the nature of the biological world, what the “pushmi-pullyu” organisms are caused to do would sometimes be morally repugnant. Insofar as human organisms have not detached from the causes operative in that biological world, they will likely be caused to do all sorts of morally

repugnant things which have arisen as our proper functions through the diachronic causal skein that has given rise to us.

Human beings, however, are not, at least not strictly, “pushmi-pullyu” organisms (recall, however, that one of Millikan’s huge struggles is to make any sort of headway explaining how this could be so, given materialism). Theorists attracted to evolutionary accounts of morality seem to have a selective memory when it comes to this fact. When dealing with bad “norms,” like hitting a rival to gain access to a female, the fact that humans are not “pushmi-pullyu” organisms will come through very clearly, whether as a fact/value divide or in some other form. But this does not come through so clearly when something resembling altruism or the more palatable features of a “moral brain” is being discussed.

The “detachment” of imperatives from indicatives that would allow us to find no reason to act on causal drives pushing us to hit someone with knuckles whose proper function is to do just that, is the same “detachment” that would allow us to find no reason to succumb to the push of some part of our “moral brain” to do the nice “tit” in return for the nice “tat” I received—especially if we see that doing so will be costly and it is calculated that it is perfectly viable not to (perhaps this person’s cooperation is no longer needed for the success of my projects and no one whose cooperation is needed will care). The “detachment,” in short, is a double edged sword.

So, on the other hand, to the degree that imperatives do become detached from imperatives, to that degree do we find ourselves no longer guided by the “bionorms” that have (somehow) fallen away—and the good fall away with the bad. This is where Sartre’s “existence precedes essence” and “condemnation” to freedom come in. It might also be the reason why Heidegger never attempted a moral theory (and is perhaps also the provenance
of his own moral recklessness). To the extent that the “detachment” is realized, there is no external guidance; there is only freedom with respect to the external world. This also is where Korsgaard’s Kantian moral theory enters in, spinning (somehow) in isolation from the causal world of selective history. I have already given my reasons for not finding Korsgaard’s way forward hopeful, but the point here is not whether such a path offers a way forward or not. The point is that the path is not one that looks to the “bionorms,” “proper” functions, and such that Millikan attempts to locate in evolutionary history. In short, in the “detachment” of imperatives from indicatives everything “falls away”—both the good and the bad—and either there is no direction at all, or we (somehow) find that we are able to give ourselves our own binding norms apart from Darwin’s biological world. Perhaps the latter path will work (I don’t think so, and I have argued that Korsgaard’s prominent version of the attempt does not work), but whether or not it does, it is not a path that looks to biology for help in moral philosophy.

The huge question for the materialist is how one “pushmi-pullyu” organism came to rise above the indicativity of the “bionorms” (and this is what they are, indicative). Just how far Millikan is from answering that question was addressed in Section II. The problem for applying any of that to morality is that, assuming that some answer has been given about how we have risen above external constraint, once we have risen above it, how are we and our reasons tethered? From there a slew of questions follow quickly: What constrains us? When we make a moral mistake, what do we get wrong? What picks out one action as the right thing to do? Why do it if it is costly and the alternative course of action is viable? Why should morality trump survival or the material flourishing of my progeny? Why should morality trump the projects I’m pursuing, or my convenience, for that matter? Perhaps all
of these questions and many I have left out all have good answers, but to whatever degree
the imperatives have come detached from the indicatives for an organism, the answer will
not come from outside the organism itself.

C. Is Moral Misrepresentation Possible?

As we have seen, one of the deepest struggles for materialist accounts of
representational content in general, and for teleosemantic accounts in particular, is how to
account for misrepresentation. How is it that I can be mistaken about seeing a cat, or about
it being Gandalf who killed the Lord of the Nazgûl? It is not at all clear that Millikan
succeeds in meeting the challenge. But what about moral misrepresentation? Might
teleosemantics allow for a convincing account of moral error, or would there be a daunting
hurdle here for someone seeking to use Millikan’s theory in moral philosophy?

I begin simply by noting that if the account of moral error were riding piggyback or
were in some way dependent on the account of representational error (as seems likely), any
problems in the latter account would affect the former as well.

Second, an account of moral error needs a more robust normativity than what
Millikan might (at best) provide through her notion of “proper” functions and quasi-
normativity, and anyone seeking to make moral use of her account how the diachronic
casual skein gives rise to proper functions will need to supply a more robust normativity
somehow. Something additional will need to be brought to the various proper functions to
determine whether it is right to, say, beat on someone with knuckles to gain access to a
female, even if that is their proper function as determined by their selective history. How
could acting on that “proper” function be a moral error? Merely causal or prudential
considerations make no headway; if we say it was “wrong” to hit him because it turned out that he had a big friend shooting pool over at the other side of the bar, that is obviously a long way from moral error. The same thing is true for holding that his error consisted in failing to realize that refraining from employing the knuckles and commencing the proper function of ululations at the karaoke machine would have done the trick.

Nor does it help if the “error” is that the behavior arose from some proper function that is at cross-“purposes” with the “moral brain” or some such thing. One might easily slip into the error of thinking that the “moral brain” has the authority to trump other “proper” functions, and that when the dictates of (that part of the brain that is) the moral brain are not followed there is a moral error. When the moral brain wins out, that is “right.” Indeed. But teleosemantics does not have within itself the resources to underwrite this. For, as far as Millikan’s own theory goes (which, again, she does not apply to moral philosophy specifically), there is no reason why it should win out, and where it does, it does so not because it should, but because it has in this instance been caused to exercise the greater force. At whatever point the “moral brain” (the specifics of its function do not matter here; specify them as you like) does not win out, and some egocentric or cruel behavior carries the day, nothing has gone wrong. The causal history of the organism was such that the “moral brain” simply did not have the force to overcome the countervailing forces. And if, in our future selective history, the abilities of the “moral brain” to trump selfish and cruel behavior are radically diminished due to the reproductive success of more ruthless and self-centered people, nothing will have gone wrong and the result would not be a decline. The constellations of causal forces would be aligned differently. As Elliott Sober and David Sloan Wilson put it, “an explanation of why human beings hold a moral principle is not, in itself, a
justification (or a refutation) of that principle.”143 The “moral brain” might explain some of our behavior, but this causal account (if it is true) of the proper functioning of the brain does not show that the behavior thus explained is right. Nor does it show that behavior out of accord with the proper function of a moral brain and flowing out of other causes is wrong. Rather, a materialist explanation simply removes what is thus causally explained from the realm of reasons altogether, leaving it to be reinstated on some other grounds, if at all.

So, the question remains: What, in short, would make what the Hutu did to the Tutsis wrong, even if the behavior arose from the selective advantage similar propensities to act in that manner conferred on their ancestors in times of competition for scarce resources, and even if the behavior did in fact end up securing reproductive advantages (as perhaps it did, for all I know).144 It’s at least possible that the Hutu were performing a proper function in Millikan’s sense. If there is to be any substance to moral error, something has to overturn the quasi-normativity of a number of proper functions and render them wrong (in error, mistaken) even where they owe their existence to their having caused selective advantages and may in fact do so in the present if they are acted upon. Something must have the authority to trump Millikan’s “proper” functions if there is going to be an account of moral error; and the “proper” functions and “normativity” in Millikan’s account would seem to be at the center of the problem, rather than the solution. Granted Millikan’s account, some of


144 I am not, of course, saying that this story of selective advantage is true; but I see no way to antecedently rule it out (or some such similar story in a different situation). Nor, obviously, am I saying that the Hutu should have done this, nor that Post or Millikan or anyone else would say that if there were a proper biological function that the Hutu should have acted in accordance with it.
our evolutionary “proper” functions are what need to be shown to be wrong and merely counterbalancing them with other “proper” functions makes no headway at all on that task.

A materialist metaphysics of morals that accepts Millikan’s account of the mechanisms that produce proper functions will need to find a way of drawing normative distinctions between functions that have the same biological provenance and are equally “proper”—while appealing only to the materialist world somehow to draw the needed distinction. One biologically produced proper function will have to stand rightfully over against other proper functions standing on the same ground and show them to be wrong. This, of course, will not be an easy task.

D. “Meaning Rationalism” and Moral Reasons

Millikan makes her hostility to “meaning rationalism” plain, and the basic direction of her work is to take meaning “out of the head” and put it “in the world.” For Millikan there is a radically non-person-centered locus of meaning: it is not persons who mean things, but their evolutionary history that means through them. As she puts it, “Reasoning… is done in the world, not in one’s head,” and what we mean is a matter of “theoretical discovery.”\textsuperscript{145} Putting it mildly, it is difficult to see what place reasons might have if Millikan is right about “meaning rationalism.” \textit{A fortiori}, there is no room for moral reasons within Millikan’s system. Because Millikan lacks a good account of reasons generally, and moral reasons specifically, moral reasons arbitrariness will imperil attempts to use Millikan’s work to ground a metaphysics of morals.

\textsuperscript{145} See note 19 above.
The problem Millikan’s work creates for reasons generally should be clear at this point. Millikan uses her notion of “sense” to displace intensions (with an ‘s’) from the center of meaning, the idea of sense in turn relying on “mapping functions” that are laden with Millikan’s brand of “normativity” because of their selective history. But intensions are intimately linked with conscious reasons for holding that this particular thing flying around is a bee or that this particular act was a pious act. Recall that Socrates wanted intensional definitions so that he could know what to do; with an intensional definition of piety, he would know what he had reasons to do. Calling \( \varphi \) an act of piety and acting accordingly is justified by the intensional definition that one is able to apply consciously to \( \varphi \). That there are some connections between intensions, reasons, and normativity is clear. And because normativity does not fit easily into materialism, Millikan attacks any notion of normativity as \textit{sui generis}—and with it intensions and reasons. Meaning, intensions, normativity, and reasons all find themselves in the line of fire in Millikan’s attack on “meaning rationalism.”

One of Millikan’s clearest expressions of the basic drive of her program is worth some closer attention: “Intentionality is… divorced from rationality, as sense is divorced from intension.”\textsuperscript{146} We have just seen the reasons for Millikan’s determination to drive a wedge between “sense” (meaning) and intension. It is no less important to Millikan’s materialist account of intentionality that intentionality not be beholden to rationality. Her’s is a radically externalist program. \textit{Aboutness} is not something in the head; merely thinking that you are thinking about something does not make it so. The diachronic causal skein that comprises your selective history does \textit{that}, and it’s not something that is made so by any reason in your head. What your thoughts are about is a matter of theoretical investigation;

\textsuperscript{146} Millikan, \textit{LTOBC}, 12.
in what your purposes consist must be discovered.\textsuperscript{147} Genuine reasons have to be cashed out by theoretical investigation and are strictly subservient to the diachronic causal skein that is susceptible to such investigations. Reasons have no independent status; \textit{causes} that are in no way responsive to \textit{reasons} lead them around by a hook in the nose.\textsuperscript{148}

Reasons, then, come \textit{after} the fact. “Inference,” according to Millikan, involves acts of \textit{identifying} what thoughts are about.\textsuperscript{149} Or, again, “the core of practical reasoning processes is a \textit{search} for proof... in practical reasoning you begin with something you would like to do or to have done and then attempt to construct something like a proof.”\textsuperscript{150} The justification of the conclusion has already happened independently of the reasoning; the latter comes in on the back side as a rationalization of what has been done. By the time the rationalizing is underway the justification is a \textit{fait accompli}, having happened in the selective history of the mechanisms that led to the behavior. Reasons do not carry the justificatory load.\textsuperscript{151} As Millikan summarizes a possible objection to her program, “if \textit{something else than the way I justify my assertions} or the way they are causally derived determined my meaning, how on

\textsuperscript{147} See note 19, above.

\textsuperscript{148} Here it is interesting to note the fairly obvious problems Millikan would have in trying to meet the second of Gibbard’s “objective pretentions” of morality, namely, the feeling that when I accept something I do so for principled, rational reasons, rather than because of biological or environmental happenstance (see Chapter 3 above). It’s pretty clear that Millikan would just have to bite the bullet on that (which, given her willingness to attack “meaning rationalism,” would likely not daunt her.)

\textsuperscript{149} Millikan, \textit{LTOBC}, 13; emphasis added.

\textsuperscript{150} Millikan, \textit{Varieties of Meaning}, 205; emphasis in original. I have omitted a typographical error in this passage. In the lead-in to the above this passage Millikan states: “Practical reasoning is often described as reasoning in something like the form of a proof: \textit{I desire A; doing B will probably lead to A; therefore I will do B}. But being more careful, that is not the way practical reasoning generally goes, but only the way practical conclusions are justified to other people,” (204-5). This is highly debatable.

\textsuperscript{151} Compare Rüdiger Bittner’s ruminations about reasons as summarized in Chapter 2.
earth could I ever grasp my own meaning—know what I or the sentences I used meant? In the first italicized portion of this passage Millikan is giving an accurate statement of what her program does. Justifying meaning with reasons is irrelevant to the actual justifying grounds, which lie in an evolutionary history quite apart from the head. And we have seen that she is willing to bite the bullet on the implications of that program: you don’t know what you or the sentences you use mean (apart from a theoretical investigation of them). In Millikan’s hands, reasons are radically displaced by a causally grounded quasi-normativity. If my arguments in Section II are correct, this is just to say that causes displace reasons. Either way, reasons have no sui generis status apart from the selective history that makes them what they are.

Ironically enough for a student of Wilfrid Sellars (though perhaps more consistently than her teacher), Millikan has no place for “the logical space of reasons.” As Christopher Tollefsen put it in a helpful essay, “What characterizes the logical space of reasons, by contrast with the logical space of nature, is that it is a space of judging—a space in which we are sensitive to reasons in light of which we exercise our freedom to judge responsibly.” Tollefsen argues that essential to an ethics that is “above the line,” that is, an ethics operating in the logical space of reasons, is “the notion of ‘rationally relevant considerations.’ Such considerations must be available to an agent contemplating action, and they must be considerations bearing upon the possibility of action.” (68). It’s safe to say that Tollefsen and Millikan would have some significant disagreements. Millikan would subordinate the “rationally relevant considerations” to all manner of forces that Tollefsen

152 Millikan, LTOBC, 10; first emphasis added; second emphasis in the original.

could only view as irrelevant (at best) or vitiating to the practical reasoning essential for ethics.\textsuperscript{154} Moreover, where Tollefsen takes reasons to be important as “considerations bearing on the possibility of action,” Millikan has reasons and intensions coming in only on the backside as rationalizations for what has been justified by an evolutionary world, and not by a reasoning process in one’s head. For Millikan, just about anything of interest is going on “below the line,” and any attempt to build up moral norms from nature by means of her notion of the production of proper functions by the diachronic causal skein of evolutionary forces will need to deal with questions of whether something “above the line” can be made compatible with the central tenets of her program and, if not, whether we still have something that can rightly be called moral philosophy. It should also be clear that a good deal of work would need to be done to show just what has authority over the “proper” functions that are orthogonal to morality—and how whatever that might be is compatible with materialism.

It would appear that Millikan’s philosophy of representational content and her story of the etiological production of the quasi-normativity at its heart is a rather unpromising place to look for a groundwork of a “metaphysics of morals.”

\textsuperscript{154} Kant would see “pathology” all over the place in Millikan when pushed into the moral realm. It’s not clear that he would need to object to her points in the theoretical realm; indeed, one of my questions for Kantians is how exactly practical reasoning is supposed to come apart from a theoretical reasoning that aspires—as Millikan’s own theory does—to explain everything, including all of the mental, in causal terms. Where, finally, is there room for the noumenal self that can be responsive to reasons as such, and how does it get traction in the theoretical world as laid out by theories like Millikan’s (the transition problem)? I have argued that Korsgaard has no good solution to these kinds of questions.
IV. Post’s Attempt to Move from Nature to Norms

Along with Ruth Garrett Millikan, David Papineau is one of the leading teleosemanticists, carrying on a project quite similar to Millikan’s. Post leans heavily on Papineau’s work, especially his “standard model of scientific reduction.” Yet, as Post acknowledges, Papineau is not at all sanguine about moving from the proper functions he believes the biological world affords to the moral norms characteristic of human beings. Papineau states flatly that he doesn’t think that “teleosemantics itself says anything about norms, nor therefore anything about how to reconcile normativity with naturalism.” In a statement Post quotes in part, Papineau continues on to give a central reason for being skeptical of attempts to transmute one into the other: “Whatever norms are, I take it that they must involve some kind of prescription, some kind of implication about what ought to be done. This simply isn’t true of the biological facts on which the teleosemantic theory rests.” Papineau puts his finger on a key problem: The facts of biology don’t prescribe anything. If Papineau is right, the attempt to press teleosemantics into the service of a metaphysics of morals is simply misguided because the biological facts just don’t underwrite the prescriptivity of normativity.

---

155 Papineau, “Normativity and Judgement,” 21, note 5.

156 Papineau, “Normativity and Judgement,” 21, note 5. Post quotes this in From Nature to Norm, 41, although he modifies the quotation slightly. In this note Papineau makes it plain that this is a difference between his account and Millikan’s. Papineau is not willing to say that the “biological design” he argues for has a “normative” element; at the same time, it’s difficult to understand what could be meant by talking about a teleology with no normative element. It does seem that something that has truly been designed would have something that it ought, in some sense, to do. But, perhaps, as Fodor argues, this points to difficulties in “adaptationist” readings of Darwinian biology.

157 Papineau himself endorses a kind of “error theory” of morality in Papineau, Philosophical Naturalism. It’s not clear whether and to what extent Papineau is retreating from his earlier stance in “Normativity and Judgment.”
Undaunted by Papineau’s observations, John Post has recently set out to build a materialist “metaphysics of morals” from Millikan’s account of the etiological production of proper functions and the quasi-normativity arising from an organism’s evolutionary diachronic causal skein. As will emerge clearly below, Post’s metaphysics of morals rests on a “bridge principle” he calls MORAL, which in turn “is a special case of bridge principle GOE.” But GOE says that an organism “ought” to do something if and only if it has an adaptation that is “circumstantially for” getting the organism to do that thing. This, of course, is straight Millikan, and, in accord with her understanding of the quasi-normativity accruing to functions by dint of their etiology, Post affirms that “the ‘ought’ here [in GOE] is a biological ought, and thin.” So, if Post is going to lay down a metaphysics of morals on the biological groundwork Millikan has laid, he will need to show how he moves from the thin, biological “oughts” of Millikan’s etiological account to the authoritative prescriptivity of moral normativity. What is it about one cluster of materialist evolutionary forces that makes it right for it to override another cluster of such forces, such that a person acting in accord with these forces is thereby morally right? Something like that is needed if we are to move from nature to norm setting out from an etiological account of proper functions and the “thin” or quasi-normativity (putatively) joined with them.

Post’s profound debt to Millikan raises questions about how much Millikan’s difficulties become Post’s difficulties. Post’s project is not to defend Millikan from the

---

158 Post, From Nature to Norm, 131.

159 Post, From Nature to Norm, 100.

160 Post, From Nature to Norm, 100. According to Post, a thin ought is an ought that applies even to thermostats (“the thermostat ought to be keeping the house warmer than that”). Compare Millikan’s thoughts about adrenaline. More will be said about the “thin sense” of ought below.
serious criticisms critics have leveled against her own project with language and intentionality, so it is difficult to say just how he would respond to them. What is clear is that he has not responded to them. So, to the extent that the problems dogging Millikan’s project affect Post’s moral project, Post does nothing to solve them and those problems devolve upon his project. At the same time, Post will likely face many of the problems noted above regarding attempts to apply Millikan’s thought to moral philosophy. Before turning to consider the specifics of Post’s attempt to move from nature to authoritatively prescriptive moral norms, two brief observations about Post’s project in light of the preceding examination of Millikan’s work are in order.

First, Post’s strategy of pressing from below is only as good as Millikan’s account of the etiological production of “bionorms” that undergirds it. While Post does not take Millikan’s “bionorms” directly over into moral philosophy, he does take over her account the production of that quasi-normativity. If the arguments that I and others have made against Millikan’s account of how “normativity” arises from a diachronic causal skein are correct, the very “normativity” at the heart of Post’s strategy of “pressing from below” is deeply flawed and cannot support the weight Post needs it to bear. Post’s project would be vitiated from the outset.

Second, it should be clear from the critique of Millikan, and especially section III above where we saw some potential difficulties facing attempts to press Millikan’s account of the production of quasi-normativity into service of moral philosophy, that Post has set himself a very difficult task. I will not review all of these problems here, but merely remind us of a few of the difficulties. Millikan herself is clear that her account does not have any prescriptive dimensions, and it is questionable whether we have any sort of normativity in
Millikan at all. Furthermore, as Post acknowledges, Millikan-type “proper” functions are all over the place when considered morally—hitting a rival with knuckles (or whatever) is no less a “proper” function than exercising a tit-for-tat “altruism.” This leads to several problems for Post. Proper functions are part of the problem; what needs to be overturned are precisely some of the proper functions that are also to be the ground of morality. Post needs some proper functions to come against other proper functions and rightfully overturn them. There will need to be some account of the authority of some proper functions (not merely their greater force) to trump other functions that are no less proper in the biological sense. An example makes this clear: What is left of morality if men turn out to have a proper function of abusing their step children, as some sociobiologists argue, and that is strong enough to trump countervailing forces? A moral theory worth its salt has to say that the proper function with less force still has authority. Some convincing account of moral authority and error needs to be given. Additionally, if Post assumes the detachment of human beings from their biological facticity (the “pushmi” and the “pullyu” come apart), then it is not clear how any biological norm could have authority over us, nor why we who are able to aim explicitly at our survival or the flourishing of our projects in the wake of the detachment should not aim at those things above doing what is moral when it is deemed (often correctly) viable to do so. Moreover, Millikan is clear that her attack on meaning rationalism is an implication of her view of the production of proper functions. Post would need either to show that she is wrong about this, or else show that the fact that moral justification goes on in the diachronic causal skein of the biological world rather than in our heads is not a problem for morality. That is, either he would have to argue against Millikan’s etiological account of the production of proper functions, or else he would have
to devise a metaphysic of morals that is “below the line” (that is, where moral reasons are not central). Finally, if Post appeals to justification, he would need to reckon with Millikan’s own views on justification being a _fait accompli_ by the time we get around to retrospectively working through it in practical reasoning.

This is a daunting catalog of problems. Some of these problems are not addressed by Post at all, and in what follows I will not be pressing them.\(^\text{161}\) The main point is clear, however. In short, if you want to move from nature to norm using Millikan’s notion of the production of quasi-normativity by a diachronic causal skein of materialist evolution, you’ve got a _very_ long way to go (as was made clear in the critique of Millikan, especially in section III).

But here I propose to lay such problems aside except as they come up directly in relation to Post’s _own_ attempts to construct a materialist “metaphysics of morals” from out of Millikan’s thought about “proper functions.” The aim of this section is to examine Post’s project within the limits of what it seeks to accomplish and to make plain the serious difficulties that remain. In the end, I will argue, Post’s “metaphysics of morals” has not offered a convincing solution to the problem of “moral reasons arbitrariness”—and none seems to be on the horizon down the road he is traveling.

### A. Paradox and “Pressing from Below”

Post’s philosophical work has been deeply marked by a desire to formulate a “non-reductive” physicalism that holds together things that seem incompatible on the face of things. Post seems to relish an air of paradox, even stating near the beginning of his most

\(^{161}\) This is because of issues of space, not because I don’t think that these are serious problems for Post.
important book that “what lies ahead is an extended exercise in having one’s cake and eating it too.”\textsuperscript{162} And in a more recent essay that is a forerunner of \textit{From Nature to Norm}, Post states that one criterion that an adequate materialist theory should meet is the “craziness condition,” going on to say he is “confident you’ll come to agree that [his] theory satisfies it.”\textsuperscript{163} By this Post means to emphasize just how revisionary his project is and how radically he departs from the ways most philosophers have carried on their work. In this Post is very much like Millikan, who also trumpets her departure from “the tradition” as a breath of fresh air in a philosophical discourse dominated by a moribund methodology that has run its course. For Post, as for Millikan (witness her attack on “meaning rationalism”), paradox is a byproduct of an original methodology that takes materialism seriously.

But for all Post’s willingness to entertain paradox, he is aware that some aspects of attempting to move from biology to moral normativity strike many as plainly implausible and that he needs to soften the paradoxes if he is to persuade anyone of his position. Not everyone thinks you can have your cake and eat it too.

One of the chief places where Post needs to soften the paradoxes in order to make his view plausible concerns what would seem to be the all or nothing nature of the move from an etiological account of the production of biological “normativity” to actual, prescriptive morality that addresses persons with reasons why they ought to act in a certain manner or believe a certain thing. In short, while it might be the case that human

\textsuperscript{162} John F. Post, \textit{The Faces of Existence: An Essay in Nonreductive Metaphysics} (Ithaca, NY: Cornell University Press, 1987), 17. In this book Post attempted a rather different way of trying to fit normativity together with physicalism than the one he attempts in his most recent work. I shall be concerned only with his most recent attempt.

evolutionary history has caused us to follow a kind of tit-for-tat “altruism” in some circumstances, it is undoubtedly the case that the biological diachronic causal skein would have caused other, not so savory behaviors. Perhaps human males do have a disposition to hit rivals with their knuckles, the relevant portions of the brain and the hand thus being “for” that. It has variously been argued that men have dispositions to adultery, that there are biologically caused dispositions to abuse one’s stepchildren, and that human animals are biologically hardwired to discriminate against members of an “out-group.” And the linguistic and social mechanisms that are “relationally for” (Post speaks of them being “RFOR”) getting us to cooperate with other human animals whose assistance would benefit our projects will cause a wealthy human who grew up in a Unitarian group in Boston in the 1930’s to act quite differently than the human who grew up in a disenfranchised merchant class in the Germany during the same period. The very different “derivative proper functions” (Post speaks of being “circumstantially for” or “CFOR”) of their relational mechanisms (think, again, of the chameleon) would have the same status as “proper” functions on these biological grounds. In short, on this account it seems, as we saw with Millikan, that whatever has been caused to be through the biological diachronic causal skein should be exactly what it is; biologically speaking, whatever is in accord with an organism’s history should be. There is a problem of how to work “misrepresentation” into the causal account.

The problem facing Post is how to have the cake and eat it too: How will the diachronic causal skein of materialist evolution underwrite moral norms where people are addressed with authoritatively prescriptive moral reasons, while not underwriting all

---

164 I do not mean to say that any of these particular claims are true; but surely it would be pure dogmatism (and just very unlikely) to claim that human animals that evolved in the manner of Darwin’s theory would not have any aspects of their nature that were morally repugnant.
dispositions that share the same biological provenance? Which ones have the requisite prescriptive authority and which do not? Unless a principled difference can be specified between the cases—something more than a “stance” the theorist brings to the diachronic causal skein—tit-for-tat “altruism” has no better claim to prescriptive authority than does beating a rival with one’s knuckles. Either the diachronic causal skein underwrites the authority of both or it underwrites the authority of neither. Post needs to show that you can have the one without the other somehow.

But the problems go even deeper; for how could the biological diachronic causal skein address any prescriptive reasons at all to persons? This is Papineau’s question above, and it is also something we have explored in relation to Millikan’s work. Here too it looks like there is an eating and having problem. On the one hand, as a physicalist, Post wants his story of normativity to be a respectable story of efficient causality operating diachronically. Thus, when faced with a criticism that the normative properties he is arguing for are merely epiphenomenal, Post responds by holding that what “‘grounds’ this [normative] explanation—but makes it no less an explanation—is a longer one in the background”—an explanation that, as Post makes clear, is of the materialistically respectable causal type.\footnote{Post, From Nature to Norm, 78.} The real work is being done at the physical causal level, and the higher level explanation in terms of normativity is materialistically respectable because it can be cashed out in the causal terms for which it is standing in as a kind of shorthand.\footnote{It looks a great deal like this shorthand and its “normativity” owes to a stance brought to the diachronic causal skein. More on this below.} At the same time, Post needs this same diachronic skein of efficient causality to underwrite the authoritative prescriptivity of moral reasons addressed to persons. So, while the putative normativity is a way of viewing
the diachronic causal skein that stands in the background and “grounds” the norms, it must at the same time really preserve the prescriptive authority of moral reasons addressed to persons in its own (authoritative and seemingly *sui generis*) terms.\(^{167}\) We see the diachronic causal skein operating exactly as we would expect: a heart pumps blood because it has been *caused* to do so by the lengthy chain of ancestors who outran predators (or some such story) because their hearts pumped blood; so also a guppy is *caused* to behave in accord with a form of tit-for-tat altruism because its forebears survived at a slightly higher rate by being disposed to behave in ways increasingly approximating tit-for-tat.\(^{168}\) But what we do not encounter anywhere in this diachronic causal skein is independent, authoritatively prescriptive reasons addressed to persons, directing them to act or believe in a certain way. The causes prescribe nothing, have no authority, and do not address persons at all. They simply cause things to happen. How some of the causes in the seamless diachronic causal skein either become or give rise to independent, authoritatively prescriptive norms is quite unclear on the surface of things. For a materialist like Post, such authoritative prescriptivity cannot be there from the start, pre-existing the operation of evolutionary forces in some Platonic way. So Post will need to identify some natural (that is, materialist) thing that happens along the way that introduces it.

Faced with arguments that normativity must be authoritatively prescriptive or points about not hitting rivals with knuckles biologically “for” doing so or, Post has a stock

\(^{167}\) Or at least any “metaphysics of morals” that would not succumb to moral reasons arbitrariness must do so. Post makes it clear that he wishes to capture the authority requisite to moral normativity; he does not take himself to be advancing an eliminativist position with respect to the authoritative prescriptivity inherent in moral norms.

\(^{168}\) *Post, From Nature to Norm*, 116-18.
response in *From Nature to Norm*: “Yes, of course.”

Post thus acknowledges the force of the points, while asserting that they can be incorporated into his view. Faced with these problems and the paradoxes they seem to give rise to, Post thinks he has a way of softening the hard edges of the paradoxes and answering tough questions like those above.

At the center of his answer is a methodological approach he calls “pressing from below.” As Post puts it, “the idea is to start with something much simpler and more tractable than moral normativity, in order to challenge certain orthodoxies about the place of normativity in the world, [and] then see what implications, if any, this approach might have for other kinds of normativity.”

Starting with Millikan-type “bionormativity,” Post tries to defang a number of objections to moving from biological facts to normativity that many have thought definitive, from Hume’s fact/value divide, to open question arguments, to Mackie-style “queerness” objections—and much else. These are the “entrenched objections” to any attempt to move from nature to norm. Post’s idea is to start with a simple, more tractable normativity—the kind arising from Millikan’s etiological account of proper functions—and see if he can make some headway on these standard objections. Perhaps if there is a basic and clearly natural normativity that is genuinely normative, it will show a way to deal with these objections that can be applied to higher forms of epistemological and moral normativity where there is a authoritative prescriptivity inhering

---

169 *Post, From Nature to Norm*, 41, 42, 49, 134, and 140. One does start to feel that there should be a limit on the “Yes, of course,” response. Post sometimes swallows down rather large mouthfuls thusly; but not everything can be swallowed down that easily. I do not mean to imply that Post’s responses consist of nothing more than this, but he does swallow quite a bit in this manner that would seem to be a bit more problematic than he allows. One is glad to see these admirable intuitions in place, but the tough question is justifying them within a materialist framework. In this note I merely voice a complaint that Post moves rather too quickly with objections that at times seem more serious than he allows.

in reasons addressed to a person. Perhaps along the way this authoritative prescriptivity will be seen to be tractable in materialist terms after all. In the end it turns out that Post thinks that reflecting on this “thin” biological normativity opens the door for nothing less than a metaphysics of morals.

Post is very clear, however, that moral norms cannot be inferred from “bionorms.” In fact, here is how Post begins From Nature to Norm: “There is no way to get from nature to norms, if ‘get from nature to norms’ means inferring authoritative moral norms from bare facts about nature. Indeed, facts about our biological nature can be a dangerous place to look for moral norms.”\(^{171}\) What we find in the biological world is a “thin” sense of ought. The “thin” sense of “ought” is what we find in the way people will talk about a thermostat. It is for keeping the house at a desired temperature, and it ought to regulate the temperature of the house. “This thin ‘ought’ is obviously not a moral or ethical ought; thermostats and hearts are not morally obligated to do what they’re for.”\(^{172}\) And the thin sense of “ought” allows for no inference to an authoritatively prescriptive moral norm. If there is a part of our biology that is “for” (thin “ought”) out-group discrimination, this says nothing at all about whether or not we should (authoritative ought) discriminate against an out-group. Moreover, Post acknowledges that this holds not only for the unsavory bionorms, but also for the nicer ones, like tit-for-tat or other deliverances of a “moral brain.”\(^{173}\) No inferences may move from the “normativity” of the biological world to the independent, authoritative

\(^{171}\) Post, From Nature to Norm, ix. Other places where Post makes it clear that no such “inference” from a “bionorm” to an authoritatively prescriptive moral norm include pp. xii-xv, 2, 8, 23-4, 42-3, 94-5, 128, 132, and 145.

\(^{172}\) Post, From Nature to Norm, 30.

\(^{173}\) Post, From Nature to Norm, 128-9.
prescriptivity of moral norms. The strategy of “pressing from below” will not be able to move in any direct way toward moral normativity.

Post’s strategy of “pressing from below” is no magic bullet, and it’s obvious that daunting challenges will press him as he pursues this project. As we assess Post’s efforts to make progress on moving from nature to normativity using the “pressing from below” methodology, we will do well to keep our eye on the main issue. Quite simply, we need to see how the authoritative prescriptivity is being brought into the picture and how it arises from the same diachronic skein of efficient causality that gives rise to the biological “normativity” that is orthogonal to moral norms and responds to (or is a stance taken with respect to) causal forces. How can such morally uneven causal forces that give rise to tit-for-tat “altruism” and knuckles “for” beating on rivals ever give rise to a constant and authoritative morality? Where and how, exactly, are the causes transmuted into authoritatively prescriptive moral norms or the grounds for them?

B. The Adequacy of Reductions

In attempting to move from nature to norms, Post seeks to reduce moral norms to materialistically acceptable stuff without eliminating them. In doing this he draws on what he calls the “standard model” of scientific reduction as explicated by David Papineau. While there are many questions about this way of proceeding that could be explored over the course of many pages, including questions of whether or not some of the specifics of the methodology end up begging important questions, we are going to zero in on only one issue: What makes a reduction adequate?
Post illustrates the Papineau-model of reduction by using the example of ball lightning. Ball lightning is a “folk” description (thought by some to be mythical) of floating balls of light that are red, odiferous, and hissing. The Papineau-model of reduction attempts to find a “theoretically interesting state” that accounts for this folk description by relating it to “the basic laws governing physical causation.” The method floats trial balloons—possible explanatory hypotheses of the right (materialist) sort—in looking for something that accounts for enough of the folk description to satisfactorily explain it in new terms, bringing it within the ambit of “a powerful, unifying explanatory theory” that makes clear the “underlying nature” of that which is being reduced, thus allowing us to “become better able to understand the behavior” of that which is reduced. In introducing a physical reductive base, we introduce “properties that are causally efficacious” and thus give the reduced phenomenon a real standing within the world. At the end of a successful reduction we can say that the physical base realizes that which is described in such-and-such a way by the “folk.” Thus H₂O realizes water, the latter having been successfully reduced to the former. It is also possible that there is a successful reduction in which high density plasma realizes ball lightning. Post hopes to use Millikan’s work on the etiological production of proper functions to show that there is a materialistically respectable something that realizes moral norms.

Questions leap to mind regarding the relationship between the folk description and the reductive base. Is just any reduction adequate to the folk description? How much of

---


175 Post, From Nature to Norm, 15; the quoted passages are taken from Post, who is in turn quoting Papineau, “The Status of Teleosemantics; or How to Stop Worrying about Swampman,” 282, 285, 287.
the folk description must be preserved in the reduction for it to go through? How much can a reduction flaunt the folk description before it passes into eliminativism? On the one hand, Post insists that theorists must be allowed to “put the received notion on trial,” reforming it in light of theoretical considerations.\textsuperscript{176} If one gives the folk description absolute preeminence in these matters, for result would be a \textit{de facto} conservatism in which received notions rule.\textsuperscript{177} If folk concepts of mass are not strictly preserved in Einstein’s theories, does that mean that those theories are inadequate and should be abandoned? Clearly not. Post is quite clear that reductions aim not at a conceptual analysis, but at capturing a theoretically significant (i.e., materialist) state that realizes whatever of the folk notion may be salvaged.\textsuperscript{178}

On the other hand, Post doesn’t want it to be the case that just anything goes with respect to the folk notion that is being reduced. “Does any of this mean that the new notion need bear no resemblance to the old, that there are no received constraints whatever on using the same term in the new way? Not at all.”\textsuperscript{179} There must be a clear sense in which the materialist base grounds essential parts of the folk notion, explaining them more fully and showing why the phenomenon behaves the way it does. “In general, for any theory, however revisionary, there needs to be some sufficiently strong resemblance between the new usage and the old, if the theory is to count as a theory of a phenomenon by the same

\begin{itemize}
\item[176] Post, \textit{From Nature to Norm}, 57.
\item[177] Post, \textit{From Nature to Norm}, 55.
\item[178] Post, \textit{From Nature to Norm}, 16.
\item[179] Post, \textit{From Nature to Norm}, 19.
\end{itemize}
name, or as non-eliminative.‖ At the same time, what counts as “sufficiently strong resemblance” will vary based on a number of factors, including the discipline in which the reduction is being made and the nature of the folk description. “We should not,” Post states, “expect some context-independent, across-the-board criterion for sufficient-strength-of-resemblance between the old and the new.” Matters of the adequacy of the reduction must be decided on a case by case basis. What may be a sufficiently strong resemblance in one case will not be a sufficiently strong resemblance in another. When that threshold is not reached, whatever it is, reductionism has crossed into eliminativism.

Post does not want to be an eliminativist with respect to moral norms, but instead seeks a reduction of genuine, prescriptive moral norms to a materialist base that vouchsafes their genuine efficacy and ensures their place in a view of reality that takes science seriously (i.e., materialistically). But he has not given us much to work with in terms of how we are to judge his success in the endeavor. “Sufficiently strong resemblance” that varies from field to field and case to case is not much to go on if we’re trying to decide whether a reduction succeeds or lapses into eliminativism. Indeed, one is even more at a loss as to how to judge the adequacy of a reduction when Post attacks Frank Jackson’s contention that an adequate reduction must retain “any feature that is ‘essential’ according to the folk conception.”

Apparently, then, a reduction can be adequate even when it dismisses “essential” features of the folk notion that is being reduced. This cries out for a lot of additional exposition, but

---

180 Post, From Nature to Norm, 19.

181 Post, From Nature to Norm, 19-20.

182 Post, From Nature to Norm, 20. The source for the position with which Post is disagreeing is Frank Jackson, “Responses,” Philosophy and Phenomenological Research 62 (2001): 653-664. In the line right after the passage I quote, Post dismisses Jackson’s position by saying that it would make Einstein an eliminativist about mass. This seems a bit hasty.
none is given. For example, if the water-$\text{H}_2\text{O}$ reduction had not captured, say, water’s being wet or freezing into something solid and slippery when it got cold, could any such reduction have been adequate? It seems safe to assume that Post would say “no.” But why?

Let’s push on this a little bit. Take ball lightning. Say that the physicist comes up with a reduction that is a materialist beauty. High density plasma has impeccable materialist credentials and it makes ball lightning fit into the basic laws governing physical causation, vouchsafing the causal efficacy of what the folk call “ball lightning.” The only problem is that high density plasma, let’s say, neither floats, glows, is red, hisses, nor has an odor. Has ball lightning been reduced; or, if the theorist insists that there is nothing but this for ball lightning to be, has it been eliminated? And, if the ball lightning-high density plasma reduction does not account for those properties (red, hissing, etc.), and one says to the physicist “That is all very interesting, and I see how it fits into your theory, but what I want an explanation of is that’ (pointing to glowing red stuff floating just so as it hisses and gives off a characteristic odor), hasn’t the reduction been refuted in some way? Perhaps the physicist could say, “Ah, well, yes. That isn’t ‘ball lightning’ you see.” But one might still ask for an explanation of whatever that red, glowing, hissing, smelly stuff is that’s floating over there. Or, perhaps, the more the folk think about what they thought was ball lightning in light of the proposed reduction (that has since made its way into all junior high science texts), they come to think that there wasn’t really the glowing red, hissing stuff after all. Something, then, has been eliminated. The red, glowing, hissing, smelly stuff that was called ball lightning doesn’t exist; instead we think of “ball lightning” as a blue, cloud-like, odorless

---

183 For my own part, I think some notion of “essential” features is probably on target. Of course everything will depend on how that gets played out. Below I will simply operate with some straightforward cases.
stuff that rolls along the ground like fog. Perhaps we drop the idea of “ball lightning” altogether... unless, of course, we can point to some red, odiferous, hissing balls of light.

Or, to take another example, consider a reduction of pain as the firing of C-fibers that is materialistically compelling. Say the only problem is that people don’t really seem to mind the firing of their C-fibers. Clearly such a reduction is implausible, and a theorist who clung to it would be (laughably) eliminativist. All such a theorist has done is change the subject.

Turning now to Post’s reduction of authoritatively prescriptive moral normativity to proper functions in Millikan’s sense, we encounter a problem regarding the adequacy of the reduction. Let’s be clear about the “normativity” that arises from a diachronic causal skein. As we have seen above with Millikan (whose view Post is carrying over), such “normativity” neither gives reasons, nor addresses persons, nor is used by anything capable of addressing others. It is neither to be done, nor to be pursued (if knuckles are “for” hitting, that does not mean that doing so is to be done or to be pursued).\textsuperscript{184} It is not prescriptive, has no authority, is not action-guiding, and (if my arguments are right) is not other than causality except as viewed through the eyes of humans who have a particular interest they bring to

\textsuperscript{184} Post has an account of a kind of “to-be-pursuedness” in Post, From Nature to Norm, 102-103 (see also, 46-49) that is tied to two of his “bridge principles” that are meant to effect the transition between the materialistic base and genuine normativity. ETBP states that “\(E\) has to-be-pursuedness-by-\(A\) if and only if \(A\) is (circumstantially) [Millikan’s ‘derivatively’] for \(E\),” (102). Here we need only note that in this sense hitting a rival (say, “Bubba”) with knuckles that are “for” doing so has “to-be-pursuedness” in this sense. GOEP, the second bridge principle (102) has the same consequence. Such “to-be-pursuedness” is just not what is in view with actual normativity and makes no explanatory headway toward the latter that I can see. Earlier Post points out that the sort of “to-be-pursuedness” characterizing the “primitive normativity” is “the effect of \(A\)’s past instances in virtue of which \(A\) was selected for,” (47; emphasis added). This makes clear again that hitting rivals with knuckles has \textit{this} sort of “to-be-pursuedness” and that it has it merely as an effect of a diachronic causal skein. We are very far from normativity indeed.
the diachronic causal skein (Post calls this the “what-for” stance). But what is left of normativity? The “normativity” arising in this manner is very thin indeed—so thin that one wonders if what we have is a reduction of normativity at all. It is rather like a reduction of “pain” to some physical state people don’t mind having, or like a reduction of ball lightning to something that is neither glowing, red, odiferous, or floating. One gets the same impression of Post’s exceedingly thin “normativity.” Moving from nature to that just isn’t all that interesting.

Moreover, even if there is a sort of quasi-normativity arising from Millikan-type proper functions (as sections II and III of this chapter gave reason to doubt), such “normativity” is so thin that it can’t do the work Post’s method of pressing from below needs it to do. Recall the basic idea: Post’s methodology starts from Millikan’s etiological account of the production of proper functions and tries to show that the “normativity” in view in Millikan’s theory can solve a number of the classic problems in moral philosophy that have been thought by many (perhaps most) philosophers to shut down attempts to move from nature to normativity. Post hopes to show that “the stock objections all fail in the case of the primitive normativity involved in a biological adaptation’s being for this or that.” These lessons will then be applied to “successively less primitive kinds of normativity involved in a biological adaptation’s being for this or that.” In the end, “it turns out that this less primitive biological normativity opens the way to a surprising, and

---

185 Post, From Nature to Norm, 29.

186 Post, From Nature to Norm, xii.

187 Post, From Nature to Norm, xiv.
surprisingly fruitful approach to the metaphysics of moral normativity.”¹⁸⁸ That, anyway, is what Post hopes to show.

Now, obviously Post’s methodology hinges on the initial thin normativity being substantive enough that showing how it can handle the “stock objections” gets us some mileage with the objections as they apply to moving from nature to the authoritative prescriptivity of moral normativity. For that, of course, is what the objections contend cannot be done, and that is the problem Post wants to make tractable. Obviously, if the supposed “normativity” is just diachronically arrayed efficient causality viewed from a “stance,” we shouldn’t be surprised if we can move from nature to that. Nor would having shown that these causes can be plugged into the “stock objections” and not fall prey to them make any headway toward showing the same with respect to the authoritative prescriptivity that is what the objections are actually about. The question, then, is just how interesting it is if Post plugs a thermostat-type thin normativity into the standard objections and shows that no problems arise. Should we be surprised? How much headway would Post have made toward showing that the “standard objections” pose no problem for the move from nature to authoritative prescriptivity (indeed, any prescriptivity)?

Not much, it would seem. Let’s grant the thermostat-type “thin” normativity for the sake of argument. The proper function of a heart is to pump blood; that is what it is for. It ought to pump blood. If it does not, it is defective. Now, we’re granting that this has a thin thermostat-type normativity. It’s the diachronic aspect of the natural selective causes that introduces this thin normativity. The heart caused the survival of enough forebears organisms by pumping blood that this caused enough of them to reproduce at a higher rate

¹⁸⁸ Post, From Nature to Norm, xiv.
that caused today’s organism to have its heart. As we have seen, because of this the heart has a “proper function” (at least as we look at the events). *This* is what Post plugs into the “standard objections.” Now, take what Post calls the “Hume’s-Law Objections”: “no norm can be inferred from any facts.”189 It is hardly surprising that we could infer the diachronic skein of natural selective causes—and thus a proper function—from the facts. And thus we have moved from facts to a (thermostat-type thin) ought. The facts tell us that a heart “ought” to pump blood.

Whether any of this addresses Hume’s question is another matter. Hume was surprised that to find people reciting a litany of facts and then, suddenly, “instead of the usual copulations of propositions, *is* and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*.190 And, this relationship being different than copulative is’es, he demanded an explanation of “how this new relation can be a deduction from others, which are entirely different from it.”191 Faced with a thermostat-type thin “ought” arising from a diachronic causal skein, one suspects that Hume might not have been moved much to back off his point, which has to do with a “new relation” embodied in the oughts and ought nots. It can look like Post has not spoken at all to this new relation, but has rather given us copulations of is’es. That said, one wonders if perhaps even with Post’s razor thin “oughts” Hume might still have been inclined to see the “ought” adduced to the facts of the natural selective diachronic causal skein as owing to a human propensity to view these facts in the manner of a new relation they are disposed to spread onto the

---


facts. In any case, it seems likely that Hume would not have thought his question to have been addressed.

Nor should he. As Post works through the various “standard objections,” he appeals at key junctures to Darwin’s theory of natural selection. Thus Post claims that “if running a Hume’s-Law argument against Darwin is irrelevant, so too is it irrelevant against DFOR,” this latter being Post’s most basic statement of Millikan-type proper functions. Post repeats this strategy with the Open Question Argument, Mackie’s Argument from Queerness, and the Explanatory Impotence Argument. With each problem we find Post saying at a key juncture that since the objection creates no problems for Darwin’s notion of an adaptation, it creates no problem for Millikan’s thin sense of “ought.” Perhaps not. But the very closeness of the linkage between Darwin’s theory of natural selection and the “ought” in view raises very tough question about how much progress Post is making in addressing the objections in a fruitful way. This is because Darwin’s theory is a causal theory, and, taking Hume’s question as our example again, inferring causes from facts doesn’t address Hume’s objection. In pulling “oughts” so close to the “is’es” of Darwinian natural selection, Post ends up not addressing the objections themselves. He only switches the question, making no headway in showing how questions about moral normativity are tractable.

Post insists that none of what he writes in response to the “stock objections” aims to sanction a move directly from Millikan-type thin “normativity” to the authoritative prescriptivity of genuine moral normativity. “DFOR is about this thin normativity, not

---

193 Post, *From Nature to Norm*, 45, 48, 70, 78.
moral or ethical normativity or the kind of ‘prescriptiveness’ involved in the moral or ethical. In view of the gap between the two, DFOR does not imply, or even support, an inference from one to the other.”¹⁹⁴ This is good, since if it did, we be left with authoritative prescriptions for hitting people with our knuckles (and worse). And Post is right that a (really big) gap separates the thermostat-type “ought” and the ought of moral normativity. This does indeed cut off inference from one to the other. “Such an inference,” Post rightly notes, “would equivocate between ‘ought’ in the thin sense and ‘ought’ in some moral prescriptive sense.”¹⁹⁵

But this cuts both ways. If it’s equivocation in the case of moral inference, it’s also equivocation when Post claims that showing how his thin “oughts” handle the “stock objections” makes headway with the completely different question that we have when we run the objections with moral normativity. And, of course, the “stock objections” were designed to apply to the authoritative prescriptivity of moral normativity. Successfully showing how the objections don’t create problems for a move from a materialist nature to the thin “oughts” of thermostat-“normativity” shows nothing at all regarding the objections to moving from a materialist nature to authoritatively prescriptive moral normativity. Far from showing how the “stock objections” are tractable, Post has not budged them at all. To think so is to fall prey to the same equivocation Post (rightly) sees barring the way to inferences from the proper function of knuckles to prescriptions about what one morally ought to do with them. The same problem of equivocation that bars the way to the one, bars the way to the other.

¹⁹⁴ Post, From Nature to Norm, 42; emphasis in original.

¹⁹⁵ Post, From Nature to Norm, 42.
Post’s reduction of normativity to Millikan-type proper functions doesn’t help him make headway with the “stock objections.” What is more, if we could point (as we might with “ball lightning”) to something authoritatively prescriptive that addresses persons with reasons to act in certain ways such that they think they ought to be done and are to be pursued, then we could say that however we should understand Post’s thin sense of “normativity,” it isn’t anything like that authoritatively prescriptive, reason-giving stuff that we’re interested in—whatever we call it (perhaps we should give it squatter’s rights on the word normativity and insist that Millikan and Post look for another, less confusing and evocative term). Indeed, I will try to point to one clear example of authoritatively prescriptive moral reasons—the case of morals reasons to reject slavery—and argue that this is quite different from the “normativity” on offer from Post. Interestingly, Post does acknowledge this very different notion. As was seen above, he says “Yes, of course” to a number of points that might have been thought to pose problems for his theory. In this way Post makes it very clear that he is not meaning to be eliminativist about moral normativity. Whatever is going on with the “proper” function of knuckles being for hitting rivals, of course we shouldn’t follow such “norms.” Agreed. But it’s that shouldn’t that is of interest. That’s the normativity that needs to be shown, somehow, to arise from nature. And it’s not at all clear what the thin “normativity” that is the starting point for Post’s strategy of pressing from below has to do with that. What is more, if Post doesn’t have more to offer than a themostat-type thin “normativity” that needs the scare quotes or the “quasi-” or the schnormativity no less than Millikan’s “bionorms,” then what he ends up talking about is just too far removed from normativity to be a successful reduction of
normativity. If normativity is nothing more than that, perhaps we have an eliminativist view after all.

C. Is There a “Double-standard”?

Post complains that there is a “double standard” when we apply reductionist models to water, ball lightning, and the like but refuse to do the same with normativity. We are owed “some justification for the double standard—that such theorizing is OK for ball lightning but not for normativity.” Post argues that we allow the application of “bridge principles” which move from plasma-theoretic facts to the relevant features of the folk notion. The facts physics tells us about plasma “imply nothing by themselves about the folk role of ‘ball-lightning’ or ball lightning’s folk properties.” But when joined with suitable bridge properties, the implications from the facts of physics to the folk properties of ball lightning can be drawn. However, unless there is some relevant difference, the same move must be allowed in the case of normativity, on pain of having an ill-motivated double-standard.

Post lists twelve of these trial-balloon bridge principles in the back of From Nature to Norm, and it is these that are to carry the weight of moving from nature to normativity. As has been seen, Post insists that we do not want to infer authoritatively prescriptive norms from the thin “normativity,” since this would stick us with authoritative prescriptions for all manner of problematic actions. Responding to Papineau’s point about the knuckles, Post responds: “Yes, of course, given only the premise that X has been biologically designed to

---

196 Post, From Nature to Norm, 26. Post is arguing here against G.E. Moore’s “open question argument.”

197 Post, From Nature to Norm, 41.
Y, on cannot infer that in some sense X ought to Y. The premise must be conjoined with some suitable bridge principle.”198 Norms are not inferred from nature, but may be seen to arise from it through these biologically rooted bridge principles. Thus a biological trait’s being “for” something may be understood through the bridge principle DFOR: “A is directly for E (normative sense) if and only if E is the effect of A’s past instances in virtue of which A is selected for.”199 The higher level folk property of being for something is thus meant to be tied back to the materialistically acceptable base theory. As with Millikan, the “for” is cashed out in the causal coin of that theory.

Now it doesn’t seem to me (though some apparently see the move as question-begging200) that there is a problem with floating such a “bridge principle” as a trial balloon and seeing how far one gets in effecting the desired reduction. There does not seem to be any problem with granting Post the right to attempt to effect the desired reduction in this way. There is no “double-standard” there, at least not on my part. But of course it’s up to Post to show that the bridge principles effect the needed transition from the materialist base to the essential features of the folk notion (or, else show why these features can be dropped while still allowing an interesting sense of that which is being reduced—in this case, “normativity”). Yet Post protests that skeptics like Papineau would need to explain why the reductive model “should apply to non-normative folk roles but not to the folk-role primitive normativity involved in a biological adaptation’s being for this or that.” Such skeptics “need

198 Post, From Nature to Norm, 41.
199 Post, From Nature to Norm, 31.
200 Post, From Nature to Norm, 8-10, 25.
to justify this double standard.” Post’s remonstrations here are not entirely clear. There is no double standard in the sense that he has every permission to go ahead and try to effect the transition in the manner he is pursuing. But surely it is up to him to show that the transitions work and that the reductions are adequate and justified. The attempt to theorize in Post’s manner about normativity seems permissible; the success of those attempts is up for debate. The reductive theorist does not have carte blanche to propose just any reduction and have it go through—nor does Post think he should.

It seems that Post means to be protesting the skepticism that has greeted his attempt (and others) to get the reduction to go through. But this is not so much a double standard as it is finding the arguments unconvincing. If we allow Post to assume a bridge principle as an explanation of something “presumed to exist” and then to use it “as a premise in an extended argument in support of the self-same thesis,” it is up to him to show that that the result is an explanation that explains what needs to be explained, rather than simply eliminating it. To think that Post’s attempt fails is not to erect any sort of double standard, but rather to think that he has not given a successful reduction of normativity to nature.

There is no double standard here with respect to Post’s right to attempt the reduction he desires using his trial balloon bridge principles; but there is a well-motivated skepticism about the results of his methodology. Post seems to have changed the subject, more than he has shown how norms arise from nature. I suppose that I would be a convinced adherent of what Post calls the “Just. Not. Normativity” objection. Against folks

---

201 Post, From Nature to Norm, 92-3.

202 Post, From Nature to Norm, 25.
like myself, Post throws down a challenge: “Exactly what is it about the targeted normativity, if anything, that must be left out by any naturalist-realist reduction? On this question, my Just. Not. Normativity interlocutors have let me down. They have advanced no suggestions about what specifically is supposed to have been left out.” It’s hard to think that Post is really serious here, for many critics of positions like his have offered substantial, specific arguments regarding what has been left out by accounts like his own.

By way of review we could say the following: both “to be pursuedness” and “to be doneness” go missing, as only causes are operative in a diachronic skein and nothing is pursuing anything. There is nothing “above the line” in the space of reasons, and no persons either address or are addressed with authoritatively prescriptive reasons. Authority is missing completely; indeed, there is no prescriptivity at all, authoritative or otherwise. The very element that has struck many materialists as “queer” has simply gone missing. There is only the efficient causality of the diachronic causal chain. There are no oughts, shoulds, or duties. Nothing here overrides in a principled manner knuckles being “for” hitting rivals; and if someone does so, there is no viable account of that person somehow being in error of any sort (unless it’s having missed the fact that the big guy at the pool table is Bubba’s friend). There is only causal force to push one through over the other. I have also argued (against Millikan) that even the barest sense of “for” in view here is brushed onto the diachronic causal skein only by a stance we (somehow; that needs explanation) bring

---

203 Post, From Nature to Norm, 89. A number of the other objections Post tries to refute would also weather his attempts to rebut them, but I will not explore that here in the interest of space.

204 A quick list might include Kant and his many followers, prescriptivists of all varieties (who will think that Post has missed the key element of being directed what to do), Russ Shafer-Landau, G.E.M. Anscombe, Christopher Tollefson, Jean Hampton, J.L. Mackie (who would urge Post toward an error-theory), David Papineau, and so on.

205 See note 184 above.
to the causality operative in the biological world. To repeat: all of that is what is missing in the “normativity” at the foundation of Post’s attempts to press from below.

Moreover, there is another important difference here. Ball lightning and plasma, H\textsubscript{2}O and water are all obviously physical stuff. Unlike the case of authoritative moral prescriptions and physical stuff, there is absolutely no reason to suspect a kind-difference between ball-lightning or water and physical stuff. Of course Post’s materialism dictates that there can finally be no kind difference in the last analysis, and he is right in seeing that he must therefore seek a reduction (of at least a token-token variety). But it is the theory that is driving this. Just looking at our normative experience, there is plenty to motivate the suspicion that the case of trying to reduce normativity to material stuff might be significantly different than the attempt to reduce water to physical stuff. Water obviously just is physical stuff. This is not obvious with normativity. Post’s materialist commitments dictate that.

Faced with all of this Post could take refuge in his insistence that his project (like Millikan’s) disdains “conceptual analysis,” but is a scientific reduction “which, being revisionary, allows the folk notion to be put on trial.”\textsuperscript{206} Thus Post could declare that “being right about the folk notion is beside the point, given that the aim of the designated model of reduction is not to preserve ordinary concepts and meaning.”\textsuperscript{207} As Post states, the “aim is not to capture or conform to received concepts or usage or meaning or intuitions, including any that make up the folk role or are essential to it.”\textsuperscript{208} Post can certainly help himself to

\textsuperscript{206} Post, From Nature to Norm, 57.

\textsuperscript{207} Post, From Nature to Norm, 97. Post is referring to Searle’s skepticism (shared by Fodor and others skeptical of adaptationist interpretations of Darwin’s theory) that the idea of “selection for” captures the meaning of “for,” but rather introduces a different concept under the same name.

\textsuperscript{208} Post, From Nature to Norm, 16. Post is talking here about ball lightning. Of course, as has been pointed out, this leaves us with a huge question of the adequacy of a reduction: “I’ve reduced ball lightning to
this move, but if this is the way the wind blows, then he’ll need to get rid of his “yes, of
courses.” These “yes, of courses” are affirming the central aspects of the folk notion that his
etiologically grounded thin “normativity” fails to approach. And if he weeds out the “yes,
of courses,” he turns out not be reducing normativity to nature, but to be eliminating it in the
name of a particular theory of what nature is like. He also would be left with some fairly
unpalatable moral quandaries.

D. Justification and Authoritatively Prescriptive Moral Norms

The subtitle of From Nature to Norms promises insight into “the Metaphysics of
Morals.” Hopefully it has been made clear at this point just how very ambitious this is given
Post’s starting point. Again, the issue to keep our eye on is how the authoritative
prescriptivity of moral reasons directed to persons is introduced into the account. As Post
puts it, just before introducing his key assumption and key bridge principle: “how can we
possibly accommodate reasons and justification in such a way that a moral norm is ‘authoritative’ in the
sense of being a reason for some human activity?” If Post would avoid moral reasons
arbitrariness, this is the question that has to be answered. Only with a satisfactory answer to
this would we have a “metaphysics of morals” worthy of the name, one where morality is
“above the line” between the mechanistic space of causes and the logical space of reasons.

At the center of Post’s attempt to meet this challenge is his bridge principle

MORAL:

---

high density plasma.” “But it’s not glowing, red, floating, hissing, or smelly like this stuff over here [pointing].”
“Ah, yes, well…”

Post, From Nature to Norm, 129; emphasis added.
Moral agent $S$ morally ought to follow rule $R$ if and only if: (i) given a justified moral rule, $S$'s cognitive mechanisms are circumstantially for [Millikan’s ‘derivatively for’] (among other things) sending the rule forward to $S$'s motivating mechanisms; (ii) $S$'s motivating mechanisms are in turn circumstantially for getting $S$ to follow the forwarded rule; (iii) $R$ is a justified moral rule; and (iv) $S$ is physically able to follow $R$.  

Now this raises quite a number of questions, and, as Post notes, “the clause that really sticks out—some would say like a sore thumb—is clause (iii) that $R$ is a justified moral rule.”

But there are other questions in the background we will want to explore before examining the rather thorny problem Post has about cashing out “justification” without smuggling in the normativity that is supposed to be explained.

A first problem is that Post is working—as he must—with only biological mechanisms and it is fair to ask why the mechanism driving MORAL should trump other mechanisms for, say, pummeling rivals with one’s knuckles. After all, they all arise from the diachronic causal skein of evolutionary forces, and each function is as “proper” as the other with respect to a materialist “nature.” Of particular interest here is Post’s statement that “bridge principle MORAL is a special case of bridge principle GOE.” Looking back at GOE, one finds the bridge principle laid out this way: “An organism $O$ ‘ought’ to $E$ if and only if $O$ contains a (compound) adaptation that is (circumstantially) for getting $O$ to $E$,” where Post is clear that “the ‘ought’ here is a biological ought, and thin.”

---

210 Post, From Nature to Norm, 131; emphasis added.

211 Post, From Nature to Norm, 131.

212 Highlighting all the occurrences of the word “mechanism(s)” in the pages leading into MORAL and studying them reveals just how important mechanisms are to this account.

213 Post, From Nature to Norm, 131.

214 Post, From Nature to Norm, 100.
circumstantially for this or that is significantly less primitive than the kind of normativity involved in an adaptations’s being directly for it. 215 Certainly, however, Post can’t think that this “normativity” is any “less primitive” in the sense of being somehow prescriptive. Rather, it’s just applied more diffusely, across a wider range of situations. Thus GOE supports a “normativity” in which I “ought” here and now to hit Bubba with my knuckles because his big friend is lying drunk on the floor. Or I ought to persecute such-and-such out-group because they no longer have power to resist and doing so will unify our in-group and give us access to more scarce resources. GOE will ground all manner of proper functions, and when Post says that they are “less primitive,” we need to be clear that the scale he uses to gauge this is orthogonal to moral normativity. We have not moved any closer to the targeted normativity.

Now MORAL is an instance, albeit a “special” one, of bridge principle GOE. But so, of course, are a number of other mechanisms that are also biological adaptations, like those elements of the physiology of the hand and the brain that make knuckles “for” hitting rivals. MORAL has the same biological provenance as the “normativity” found in knuckles that were “selected for” hitting rivals. And it is initially quite unclear why MORAL should trump other mechanisms that share the same biological footing; indeed, we have to say that considered purely biologically, there is no reason why it should in all circumstances. Indeed, if the diachronic causal skein that dictates such matters causes other mechanisms to become more prominent than those at play in MORAL, it’s difficult to see what sort of adventitious something could emerge from the materialist world to trump those forces.

215 Post, From Nature to Norm, 101; emphasis in original. The connections with Millikan’s work are obvious.
Additionally, what we have on offer from Post is a story of efficient causality, at root. Different forces and mechanisms are playing themselves out. Now say that someone is caused to follow one set of mechanisms over another such that MORAL ends up not having sufficient force and the person commences pummeling Bubba with his knuckles. What would this person be getting wrong? There is no authoritative prescriptivity here (indeed, no prescriptivity at all), only mechanisms and causes. That one mechanism predominates in someone rather than another is not morally significant as such; we have only a story of how he was caused to act, rather than how he should have acted. Perhaps one might think that clause (iv) of MORAL would help because there would be no possibility of doing otherwise, given the way the forces play out. Given the way the causal forces were arrayed, the gentleman could not do otherwise than commence pummeling Bubba with his knuckles. In that case, MORAL would not be in view because of clause (iv). But it’s not clear (1) how the materialist will formulate this sense of being able to choose for or against R, or (2) what help it would be to do so. In other words, even if we grant the materialist that the “pushmis” were—somehow—to become separated out from the “pullyus” (that is, the imperative and the indicative were to become distinguished) so that the person could choose whether or not to follow R, this would not help matters. For even if the person were choosing which mechanism to follow, he would get nothing wrong in choosing one mechanism over another.\footnote{Furthermore, any such story would be leaving the attempt to move from nature to norms behind. The norms would have to arise essentially from the moment of separation itself, somehow (this again, is Korsgaard’s basic move and, I think, the basic impulse of contemporary Kantians generally).} The biology alone doesn’t answer this question. Some story beyond just biological causality and biological mechanisms would seem to be
needed to introduce an element of authoritative prescriptivity into this otherwise causal biological picture.

In any case, the first point to make is that the direction of the Millikan/Papineau/Post story of the biological world—as Papineau rightly notes—creates a presumption of moral indifference of the biological world as between the various mechanisms to which it gives rise in human animals, including those involved in MORAL. MORAL is just another adaptation, with no special claim on us. Each mechanism arises from GOE no less than the other. Absent some compelling arguments to the contrary, there would seem to be a presumption built into the foundations of Post’s view against any principled elevation of MORAL above other adaptations. Of course Post has a story here, but the initial presumption quite obviously goes in the other direction. All we have from the Darwinian point of view is two different mechanisms that are down on all fours with each other, each having sway over the next generation only as the causal forces play out one way over another. If MORAL loses sway, nothing has gone wrong.

A second point has to do with Post’s claim that MORAL is a categorical imperative, a claim he begins working toward earlier in the book. The sense of “categorical imperative” that Post employs is taken over from Mackie. Mackie puts it this way: “A categorical imperative, then, would express a reason for acting which was unconditional in the sense of not being contingent on any present desire of the agent to whose satisfaction the recommended action would contribute as a means.” After citing this in part, Post claims that an imperative central to “golden-rule altruism,” namely, “Treat the other as oneself so long as one has no basis for thinking that the other will not do the same,” is a categorical

---

imperative in this sense because it is “not an imperative that is conditional upon any present
desire of the hominid agent.”\textsuperscript{218} But the categorical imperative in Kant is rather more
stringent. Kant would not allow the following as a categorical imperative: “Do not lie \textit{so long}
as you have no basis for thinking that the other guy is lying.” It doesn’t really matter for the
obligation that you put yourself under as an autonomous agent what the other person is
doing. The obligation is not only independent of present desires but also of external
circumstances—and certainly what anyone else may or may not be doing.\textsuperscript{219} The notion of a
“categorical imperative” that Post is using is very weak.

As Post works toward MORAL, he employs this thin notion of a “categorical
imperative” to claim that what we see in “golden-rule altruism” is a significant strengthening
of what can be claimed for biological normativity; while the golden-rule imperative is not yet
a moral norm, “it is a rather strong normative ought, in the sense that it is not a hypothetical
imperative but categorical in Mackie’s sense.”\textsuperscript{220} But just how much progress has Post made
toward a genuine normativity such as found in moral norms? One, by now familiar,
problem is that due to their common etiological provenance, these “categorical imperatives”
are down on all fours with such “bionorms” as hitting a rival with knuckles “designed for”
doing so. Again, when these mechanisms are causally pitted against one another, what is to
prefer one over the other? It has already been shown that Post has significant work to do
with this sort of problem and I will not belabor the point. Second, while the “categorical

\textsuperscript{218} Post, From Nature to Norm, 120; emphasis added. This is taken from Philip Kitcher, “The
Evolution of Human Altruism,” The Journal of Philosophy 90 (1993): 506 (the page number is not entirely clear in
Post’s citation). Post first cites this passage from Kitcher in Post, From Nature to Norm, 119.

\textsuperscript{219} Nor is it clear that the categorical imperative is incompatible with all “present desires,” such as the
desire to act out of respect for the law itself.

\textsuperscript{220} Post, From Nature to Norm, 121.
imperative” in view here has already been shown to be quite weak, it would seem that even Mackie’s highly attenuated notion of categorical imperative is too strong for “golden-rule altruism.” The imperative “Treat the other as oneself so long as one has no basis for thinking that the other will not do the same” might be rephrased to bring out a very real desire at the heart of such “altruism.” Putting things crassly, we could say “Treat the other as oneself so long as there’s something in it for you.” After all, isn’t this why we’re concerned that the other play along? And if she doesn’t, then all bets are off… because there’s nothing in it for me. Clearly that is not a recognizable categorical imperative by anyone’s definition. And if the crass rephrasing is way off target, it would be interesting to hear what else it is that could be motivating the need to make sure that there is no basis for thinking that the other will not do the same. It is not clear that Post has made any progress toward the genuine normativity of moral norms with his attempt to tie “golden-rule altruism” to a “categorical imperative” of sorts.

Turning now to the supposedly “categorical” nature of MORAL, we find Post applying Mackie’s sense of a “categorical imperative” to MORAL: “In Mackie’s sense of ‘categorical,’ at least, the moral ought characterized by MORAL is categorical, not hypothetical, since the ought expresses a reason for acting which is unconditional in the sense of not being contingent upon any present desire of the agent to whose satisfaction the recommended action would contribute as a means.”221 The twist that we find with MORAL as opposed to “golden-rule altruism” is that MORAL rests upon R being a “justified moral rule.” When there is such a rule, then an agent “is morally obligated to follow R whether or

221 Post, From Nature to Norm, 133.
not doing so would contribute as a means to satisfying some present desire.” Perhaps if “justified” is suitably fleshed out MORAL may be a categorical imperative after all.

Obviously, a great deal is going to depend on what Post can make of justification. Before considering the crucial question of justification in MORAL, a few other observations should be made. We have already considered that MORAL may have some problems being a moral imperative at all, owing to its being determined by a diachronic causal skein so that (perhaps) nothing gives it a principled right of overriding other, less savory “bionorms” sharing the same etiological background. (I say “perhaps” because we have yet to explore Post’s story about justification; maybe there’s a new twist there that will make things different.) Additionally it is not clear whether Post would hold that MORAL would apply independently of whether others are holding to MORAL, so it is not clear whether or not MORAL would be categorical in Mackie’s attenuated sense.

Another observation to make with respect to MORAL is that in Post’s lead-in to MORAL he grants himself a rather large piece of the problem, namely, the moral agent herself. A rather lengthy passage that occurs twice in the pages just prior to MORAL and is repeated verbatim a third time two paragraphs after it needs to be cited in full:

we are talking about creatures cumulatively endowed with (i) language, consciousness and the (all too limited) capacity for such things as empathy, sensitivity to social (dis)approval and to other reinforcers, and a sense of fairness; (ii) reason in the form of cognitive mechanisms that are (circumstantially) for, among other things, (a) solving complex practical and theoretical problems by adopting rules conformity to which would enable a solution, as well as (b) weighing whether this or that candidate rule or solution is justified, revising or replacing the candidate if it is not, and sending it forward if it is; and (iii) mechanisms that are for motivating the creature to

---

222 Post, *From Nature to Norm*, 132-3. Note that this leaves open the question of whether others need to be committed to MORAL as well before one is obligated to hold to it.

223 Millikan, one will remember, has very little place for justification. See section III.D above.
follow the candidates send forward by mechanisms (i) and (ii). In short, suppose we are talking about moral agents.224

Post continues on from here to say that “if some endowment essential to moral agency has been left out, it can easily be added, since there will always be some (compound) mechanism that is (circumstantially) for effecting that endowment if it is real.”225 After laying all this out, Post wonders: “could the solution to our dilemma be staring us in the face?”226 Perhaps—if one could find someone willing to grant all that—it would be; but surely Fodor is gnawing on a rug somewhere.227 This really would be a lot to concede, especially the crucial part about the ability to weigh both theoretical and practical rules as to whether or not they are justified. For isn’t this precisely the normativity that is supposed to be explained?

Post argues that it is not, and here we come to the crucial issue, the one we needed to keep our eye on; for now the prescriptive authority essential to genuine normativity is in place. Where did it come from? How has Post introduced it into his theory? The third clause of MORAL lays claim to R’s being “a justified moral rule.” But how did this justification get there? It looks for all the world like Post just helped himself to it in specifying moral agency. Post proleptically addresses this objection: “Why isn’t this question-begging? After all, as Jaegwon Kim reminds us, ‘justification manifestly is normative. If a belief is justified for us, then it is permissible and reasonable, from the epistemic point of view.’” After

---

224 Post, From Nature to Norm, 130; italics in original, underlining added. The identical passage (with different emphasis) appears also on 126 and 131-2. The strong influence of Millikan is plain in this passage.

225 Post, From Nature to Norm, 130; emphasis added. The latter part of this passage reminds one somewhat of the “weak anthropic principle” in debates about cosmology.

226 Post, From Nature to Norm, 130.

227 See note 130 above.
stating the objection, Post responds that “Yes, of course, justification is normative.”

Furthermore, Post concedes that MORAL contains a genuinely normative term not only in clause (iii), but in (i) and (ii) as well; and yet he insists that the issue is “whether the term ‘justified’ is morally normative.” Post insists that “justified,” as it appears in MORAL, is not morally normative, but epistemically normative. Thus it is not smuggling in the authoritative prescriptivity of moral norms.

But this is highly problematic. To start, it’s not clear that it’s correct. The justification in view does seem to be moral. In the long, question-begging passage cited above, Post grants himself moral agency. And it certainly does appear that he has labeled that accurately. Indeed, quite a number of theorists have thought that the authoritative prescriptivity of our ability to reason, when applied to the practical realm, is the essence of moral agency. Post rightly recognizes that we have reason, and that we use this to solve “complex practical and theoretical problems,” by “weighing whether this or that candidate rule or solution is justified.” Many have thought that the authoritative prescriptivity inherent in the reality Post has described is of the essence of our ability to give and respond to authoritatively prescriptive moral norms. It can look rather like Post is explaining moral normativity by assuming moral agents who are such as to respond to moral normativity.

---


229 Post, From Nature to Norm, 134; emphasis in original.

230 Think of any one of a number of contemporary Kantian philosophers.
But let’s leave aside the fact that Post grants himself the moral agent before formulating MORAL. Are there other problems facing Post’s introduction of authoritative prescriptivity into his theory? Post’s stated way of proceeding—namely, claiming that he is only talking about epistemic normativity rather than moral normativity—is scarcely any more acceptable. To put matters baldly, Post is helping himself to a kind of authoritatively prescriptive normativity; and even if one would like to call that authoritatively prescriptive normativity “epistemic,” it is still—qua authoritatively prescriptive normativity that addresses reasons to people—exactly what needs to be explained, whatever we call it.231 A materialist explanation of authoritatively prescriptive moral normativity does not make headway by helping itself to authoritatively prescriptive epistemic normativity, for it is the authoritative prescriptivity that needs to be explained materialistically. Whether that is under the label of “moral normativity” or “epistemic normativity” simply does not matter. In the same essay Post quotes, Jaegwon Kim notes that “epistemology is a normative discipline as much as, and in the same sense as, normative ethics.”232 In short, there are shoulds and oughts that address us as reasoning persons who respond to reasons for belief. Authoritative prescriptivity is in view either way; and, to repeat, that is what needs to be explained. It does not matter in what area of life it is found or what one chooses to call it. Post has helped himself to that which needed to be explained in materialist terms, without being eliminated.

231 Millikan is much more clear about what is at stake here than Post seems to be, and she attempts to give a reductive account of epistemic normativity as well. See note 245 below.

It might occur to a materialist like Post that the difficulties might be circumvented by advocating a thin epistemology (perhaps a “naturalized epistemology” like the one Kim attacks) that is innocent of all the problematic normativity. We could let the cognitive psychologists describe what we do in fact do when we believe something.\footnote{On the other hand, witness Kim, who argues that “the concept of belief is itself an essentially normative one, and in consequence that if normativity is wholly excluded from naturalized epistemology it cannot even be thought of as being about beliefs,” (Kim, “What Is Naturalized Epistemology?” 288). Of course if “reasoning is done in the world rather than in the head,” as Millikan holds, perhaps one will be okay with that. But one might still want an explanation of why the head thinks something so very different is going on.} This would get rid of the problem, and Post’s tack of turning to the epistemic realm might be made to work. But the theorist who attempted this move would not only face the (considerable) problems attending such a position as an epistemology, but would also have removed the means by which Post is introducing authoritative prescriptivity into his materialist explanation of moral normativity. The original problem of authoritatively prescriptive moral norms that address reasons to people would be back on the materialist’s plate. Post would face the original problem in a new, epistemic register.

Remember, also, that there will be a question of the force and vigor of the causal forces of the various mechanisms that are jostling their way to the surface of our behavior. The mechanisms underlying MORAL only carry the day insofar as they are pushed forward through a Millikan-type diachronic causal chain. But one might wonder about another possible “bridge principle” that we can call VIABLE and the (readily understandable, given Darwin’s materialism) mechanisms underlying it.\footnote{Here, again, as in earlier chapters, I am employing Richard Brandt’s terminology. I think his terms are very useful for throwing a very difficult moral problem into bold relief. Unfortunately, the terminology has not been widely picked up by moral philosophers.}

**VIABLE.** Agent $S$ prudentially ought to follow a justified rule $R$ if and only if: (i) $S$ can assume others will also go along with $R$; (ii) following $R$ isn’t going to be too
costly in terms of surviving, reproducing, and carrying on one’s preferred projects; and (iii) not following R is not viable, in that (a) one is likely to be detected and needs the cooperation of the others who are likely to detect the non-compliance, and (b) one does not have the power to overcome the resistance of those who will detect non-compliance and be bothered by it.

Now Viable is offered here partly in fun—I haven’t really tried to work out a tight biconditional as a bridge principle—but only partly. After all, one need not be an unusually acute observer of human behavior to wonder if something like Viable might trump something like Moral with some frequency among humans. And the point being raised for Post is fairly straightforward: When the etiological causal mechanisms underlying these two plausible (if broadly sketched) “bridge principles” conflict in a particular agent, the diachronic causal skein of her evolutionary history having caused them to have a fairly similar causal force, there is no obvious reason in Post’s account why Moral should go forward instead of (something like) Viable. One or the other just does go forward. And if the diachronic causal forces are arrayed such that Viable usually wins out, nothing has gone wrong—except from the viewpoint of those in whom the diachronic causal skein has played out so that Moral tends to be more forceful (and those who are being trampled by behaviors carried forward under Viable, who might experience a sort of Nietzschean “ressentiment”). Indeed, given Post’s materialist world, perhaps some might even argue (I do not claim that they would be right to do so) that it is Viable that is the rational way to live, given the biological conditions we are given.

Be that as it may, what one wants is a clear answer to why the mechanisms pushing Moral should trump the mechanisms pushing Viable, even on occasions when the diachronic causal skein of our evolutionary history causes Viable to dominate. Specifically, the person who rightly judges that not following R is viable might wonder why
she should nevertheless follow R—especially when she sees that those operating according to MORAL do so because of the array of forces in their own diachronic causal skein (or maybe because they have not deemed refusal to go along with R to be viable in their situation, and so are prudentially playing along with R). 235 Again, the basic issue is how the genuinely authoritative prescriptivity of moral reasons addressed to a person in such a situation can be understood to have arisen from the diachronic causal skein of biology.

In light of all the foregoing, one might well take issue with Post when he says of the “ought” in MORAL that, “even though the ‘ought’ here, in light of GOE, is still a kind of biological ought, it is a uniquely special kind of biological ought.” 236 Post states that the biological “ought” arising in MORAL is uniquely special “partly because it is meant to apply to creatures who are endowed with” moral agency, but “above all because R is a justified moral rule.” 237 I have argued that there are a number of problems both in the background of all of this, and in the moral agency that is assumed, and in the status of MORAL itself. 238 The authoritative prescriptivity of moral normativity that was to be explained has—fairly brazenly—been smuggled into Post’s “metaphysics of morals.” And, of course, in a materialist world nothing except the theorist himself could have “meant” MORAL to apply to moral agents.

---

235 I take it that the reply is that the person does in fact know that a matter of oughts and shoulds is actually at stake and is aware of the possibility of moral error that he or she is under an obligation to avoid. The question, as I see it, is not whether that knowledge is in place (it is), but how it is to be explained—and if materialists like Post have a workable explanation for what is in fact the case.

236 Post, From Nature to Norm, 131. On GOE, see notes 159, 160, and 214 above.

237 Post, From Nature to Norm, 131, 132; emphasis added. Where I have put the words “moral agency,” Post repeats for the third time the passage quoted above at note 224.

238 As mentioned above, there are other problems that could be explored that I have passed over in the interest of not making this chapter any longer than it already is.
There are, I think, ample reasons to think Post has not given a successful materialist story of how the human animal moved from nature to norm. Having followed Millikan’s argument that a diachronic causal skein of evolutionary forces gives rise to “proper” functions and a thermostat-type thin “normativity, Post shares in the weaknesses of Millikan’s argument and never shows why the proper functions associated with MORAL should rightly override the proper functions associated with something like VIABLE when they all share the same biological provenance. In the end, however, Post ends up supporting the genuine normativity of first-order moral judgments—but only by smuggling in the independent, authoritative prescriptivity he needs.

E. When Owning Slaves Is Viable, Should One Be an Abolitionist?

Recall that in talking about a materialist reduction of “ball lightning,” Post realizes that in such reductions “there needs to be some sufficiently strong resemblance between the new usage and the old, if the theory is to count as a theory of a phenomenon by the same name, or as non-eliminative.”\(^{239}\) This is sensible enough, and of course it applies in the case of Post’s materialist reduction of moral normativity grounded in authoritatively prescriptive moral reasons addressed to people no less than it would apply to “ball lightning.” If we can point to a clear example of that which the materialist hopes to reduce, then the materialist must show how the reduction is adequate to that example—or else try to show how an eliminativism is plausible here.\(^ {240}\)

\(^{239}\) Post, *From Nature to Norm*, 19.

\(^{240}\) It’s hard to know how to take Post’s attack on Jackson’s insistence that a successful reduction must retain “essential” features of the “folk conception” other than as indicating a willingness to be an eliminativist about whatever cannot successfully be reduced to a material base. Unfortunately, Post gives his
One such clear example is the case of slavery. Again, slavery is morally wrong. The question is whether Post’s materialist reduction of moral normativity can sustain the independent, authoritative prescriptivity embodied in this clear case of a correct normative judgment: “slavery is wrong.” This has to be done without smuggling in the non-reduced moral normativity that is to be explained in the reduction.

At the heart of Post’s attempted materialist reduction is Millikan’s etiological account of the production of “proper functions” and the quasi-normativity she locates in them. As she makes clear, these “quasi-norms” are not prescriptive, and—morally speaking—their content is all over the place: given their place in the diachronic causal skein of human evolution, knuckles really are for hitting people. Particularly important for Post’s account is Millikan’s idea of “relational proper functions” and “derivative proper functions,” and Post speaks of biological mechanisms being RFOR (that is, relationally for something) and CFOR (that is, derivatively or circumstantially for something). The standard example is the chameleon: owing to the diachronic causal skein of its evolution, it has biological mechanisms that are RFOR changing its color to blend into its background. In a particular circumstance, such as sitting on a log of a particular shade of brown, its mechanism is CFOR turning that particular shade of brown so as to blend into its background in this circumstance.

Now, according to Post (as for Millikan), a human being has socially adaptive mechanisms that are “for” getting the organism to cooperate with others. “The agent is moved to treat the other as oneself when those of its mechanisms that are (circumstantially)
for motivating it to follow the golden-rule altruism rule are operating as designed and in design conditions.” 241 A portion of human evolutionary biology is “for” doing that, and human beings will tend to act accordingly in certain circumstances. In certain circumstances, then, human beings will act in accord with “golden-rule altruism” and not enslave other human beings. By the evolutionary design of human nature, when humans are in the specific conditions where a portion of our nature evolved, the history of that causal skein will indicate that part of us is “for” refusing to enslave (at least certain) others, in accord with “golden-rule altruism.” In the relevant circumstances then, when we enslave (at least certain) other human beings, we act out of accord with a portion of our evolved nature.

Of course human beings have evolved a many-faceted nature and we find ourselves in a bewildering array of circumstances. If there be such a thing as an evolved “golden-rule altruism” (I take no stand on this, one way or the other), it is one facet of our evolved nature that operates as designed in some particular circumstances with regard to at least some people. But this leaves us with a host of questions: What other facets of our nature are there? In what circumstances are those facets “designed” to operate? What are they “for” in those circumstances? What is the scope of any facet of evolved human nature, given the particular diachronic causal skein in which it arose? Is that diachronic causal skein identical for all humans with respect to a particular facet of evolved and evolving human nature (since obviously not everything is)? What if different facets of our nature come into conflict? Which should trump the other? Is the “should” even in order here, or is it a matter of which facet is more deeply rooted in a particular animal’s diachronic causal skein? Is a particular facet of evolved human nature circumstantially for being applied to all people, or, given the

241 Post, From Nature to Norm, 120.
particular diachronic causal skein that gave rise to it, does it apply only to an in-group or an out-group?

The initial point to be made here is that whether or not “slavery is wrong” is true universally, for all people, in all circumstances will hinge on quite a bit of historical, theoretical investigation. Remember, *it is the history that determines what a bit of human evolved nature is “for.”* That is the essential point in Millikan’s etiological account of proper functions, and Post is leaning heavily on this. Recall the key phrase in Millikan: “Reasoning, I insist, is done in the world, not in one’s head.” 242 The history of the diachronic causal skein that gave rise to the many facets of human nature is what matters, and whether or not slavery is wrong—at least before we get to the specifically justificatory issues of MORAL—is contingent on the results of this theoretical investigation that determines what facets of human nature are for what and what circumstances they are “designed” to operate in, and with respect to whom. As Millikan puts it, “root purposing is unexpressed purposing; our job is to discover in what this purposing consists.” 243 Whether or not “slavery is wrong” is a part of our nature, the circumstances where this applies, to which people it applies, and so on—all this must wait on the results of historical/theoretical investigation. This is hardly the ringing, authoritative prescriptivity rightly characterizing the moral judgment, “slavery is wrong.” So far, Post’s materialist reduction of moral normativity seems to be off-track.

It seems likely that the results of such investigations into evolved human nature and the “designed” application of its various facets in the many situations we find ourselves is

---

242 Millikan, *White Queen Psychology*, 12.

243 Millikan, *White Queen Psychology*, 217. A few lines later Millikan declares that her “thesis will be that the unexpressed purposes that lie behind acts of explicit purposing are biological purposes.” See notes 61-70 above.
would be a mixed bag. The list of things that “evolutionary psychologists” have speculated might be parts of that nature is long and inauspicious. Historically speaking, many human beings apparently have functioned in situations where they saw fit to enslave other human beings. In light of our inveterate historical tendencies to enslave other human beings and their continuation down to the present, the adherent of a materialist evolution has to admit that it may well be that one (perhaps powerful) part of our evolved nature is for enslaving other human beings. I don’t argue that this is the case, but merely that it would be dogmatic from the standpoint of materialist evolution to rule it out. On the contrary, it seems quite likely that with respect to slavery or other moral issues, some facets of our evolutionary nature are “for” (in the thin sense) what is clearly morally wrong. On the Millikan/Papineau/Post story, knuckles really are “for” hitting a rival for the mate you want. It has been argued that a part of the male human brain is “for” abusing stepchildren. Given our past and our present, perhaps a part of evolved human nature is “for” enslaving other human beings—perhaps those of a potentially threatening “outgroup.” Of course, as Post insists, none of this comes over directly into the moral realm of authoritative prescriptivity. But these proper functions are a part of the diachronic causal skein that has shaped human beings, and what we need is some account of how other parts of the same diachronic causal skein should rightly trump the parts that are “for” enslaving others.

If, as seems likely, there are sometimes conflicts between facets of evolved human nature, how it is that one trumps the other? Is it purely a matter of the force of some causes as set over against the strength of other causes? What if the “wrong” one forces its way to the surface? If that “wrong” facet is deeply embedded in the diachronic causal skein that gave rise to human nature and sets what it is “for,” what exactly does the person get wrong
in following what is morally wrong? Specifically, since we are moving from nature to norm, how does the biology of the situation determine that the one is right and the other is wrong, when the biology is indicating that our nature is, on balance, “for” doing what is morally wrong?

Moreover, the very mechanisms that are relationally “for” securing the cooperation of other members of the group of people relevant to the success of our projects may well serve to enforce the morally wrong behavior socially. The same mechanisms that are relationally for social cooperation, are circumstantially “for” abolition in the Boston of 1820 and circumstantially “for” slavery in the Charleston of 1820. What a facet of our nature is “for” seems dependent on circumstances in ways that make for an arbitrariness to how our biology speaks to a moral situation.

Nor does the biology of the situation speak with the requisite prescriptive authority. The diachronic causal skein central to the evolutionary story at the heart of Post’s reduction doesn’t prescribe anything to the moral agent at all, but simply causes her to do this or that. Once the verdict is in about the diachronic causal skein, there still is nothing that addresses the moral agent with any sort of prescription. We have only a story about some causes, but nothing that addresses us with reasons to do this or that as moral agents. If it turns out that strong elements of our diachronic causal skein exert causal pressure to enslave members of an outgroup, no reasons have thereby been addressed to us. And, of course, the same goes if it turns out that these biological forces exert causal pressure not to enslave others. Prescriptivity goes missing.

Authority also drops from view. What rightful claim does this biology have on you? If your knuckles are “for” hitting a rival in a situation like this where they were designed to function that way (“Bubba’s girl is really pretty and I think I can take him”), that does not
render it right to pummel him. Nor do you do something wrong if you don’t hit him. The same holds for any deliverance of evolved human nature. If your “moral brain” indicates that you should not cheat on your taxes, but you reasonably calculate that it is viable to do so and that conclude that it would be desirable to purchase the new car that could thereby be gained, you don’t get something morally wrong if all that is in view are the biological causes operative in the situation. One bit of your biology shoulders aside another bit.

The issue here snaps into focus when we recall a question we saw Korsgaard ask back in Chapter 2: “the day will come, for most of us, when what morality commands is hard … And then the question—why?—will press, and rightly so. Why should I be moral?”

Authority is in view here. It is what Ted experiences when he doesn’t want to leave a note on the windshield of the car he accidentally damaged, but recognizes the authority of the reasons for doing so. It’s what we claim to be getting right when we say that the slaveholder is wrong in her actions. The “normativity” arising from the biology and the bridge principles of Post’s theory do not give us an authoritatively prescriptive moral judgment that “slavery is wrong.”

All of these problems are solved in MORAL, of course. But the solution is by fiat. The moral agent simply appears on the scene, accoutered with language, consciousness, empathy, sensitivity to social disapproval (hopefully not circumstantially set in Charleston in 1820), a sense of fairness, and reason—complete with the capacity for “weighing whether this or that candidate rule or solution is justified.” Thus outfitted, the moral agent drives all the

---

244 Korsgaard, The Sources of Normativity, 9; Korsgaard’s emphasis. Korsgaard wants her listeners to be very clear what this question is that she is asking. To ask the question as she asks it is not, she insists, a call for an explanation of our motivations for moral behavior. Instead it is to “request an account of rightness” (9). She continues: “we are not looking merely for an explanation of moral practices. We are asking what justifies the claims that morality makes on us” (9-10). Korsgaard calls this “the normative question” (10).
preceding difficulties from the scene (or at least gives some hope that a determined full-information descriptivist or an ingenious Kantian might show that authoritatively prescriptive morality fits with materialism. Some devilish details might still complicate matters, but at least there would be hope of a way forward). Slavery is seen to be morally wrong.

One is left wondering, however, just how this is a move from nature to norms. At the crucial juncture, the moral agent simply appears, fully equipped with the unreduced, justificatory normativity that was supposed to be being reduced to something materialistic. Admittedly, it may be (as Post claims) that the normativity problem Millikan spoke of is now in an epistemic register rather than a moral register, but it is the same problem of normativity. Slavery might now be shown to be morally wrong, in the authoritatively prescriptive sense that is essential, but opening up this possibility, Post has found it necessary to leave materialism behind and appeal to a materialistically unreduced normativity. Instead of moving from nature to norms, Post has moved from moral agents to norms—and that is quite different.

Slavery, however, might be shown to be wrong, and this is more important for a moral theory than its consistency with materialism.

V. Millikan, Post, and Moral Reasons

“Reasoning, I insist, is done in the world, not in one’s head.” I doubt this. But if it were the case, then, of course, reasoning and the reasons arising from it would be whatever that world comes up with. But for Millikan, Post, and other materialists, the world that does the reasoning is a world, finally, of non-personal material stuff and efficient causality; the
“reasoning” that is done in that world is a causal matter to be theoretically investigated. Reasons are theoretical items. Even logic, according to Millikan, must either be radically demoted to something of an intellectual sideshow of interest only to historians, psychologists, and sociologists, or we must re-envision it completely, “recognizing logic as continuous with the natural sciences.”

Reasons are radically displaced; moreover, they are led around by causal forces. Our reasons, moral or epistemic, are caused to be what they are and are truly known only through investigation of the diachronic causal skein as that has played itself out. The “reasoning” of the world that is thus playing itself out in us (that has an oddly Hegelian ring) may cause us to have all manner of “reasons.” I may have reasons for pummeling “Bubba” with my knuckles, or for obeying a “golden rule” (so long as I think others are doing so), or for following MORAL or VIABLE. The reasons I have are determined outside of me by the diachronic causal skein that meets up in the animal that is “me.” Of course this diachronic causal skein will include cultural elements, such as whether I am born in Boston or Charleston in 1820, so that the reasons I am caused to have will be derived from the functioning of relational proper functions such as, perhaps, Allan Gibbard’s mechanisms of sensitivity to “conversational pressure.” Via the reasoning done in the world that has worked itself out in their respective situations, the inhabitants of Boston and Charleston will thus have very different reasons.

Much as we saw with Gibbard’s own program of trying to tie morality to the world given to us by Darwin’s theory, materialists have a difficult time finding a place for moral reasons. Millikan has not theorized about morality, but it’s not clear that she would have much to offer. Hers would be a metaphysics of morals that is “below the line” dividing the

---

245 Millikan, *LTOBC*, 274.
mechanistic space of causes from the logical space of reasons. Indeed, the driving impulse in Millikan’s work is to efface the line altogether. Nor do Post’s efforts to clamber out of the “coal pit” of moral reasons arbitrariness get him any closer to moving from a materialist nature to norms, or, rather, he gets out only by helping himself to the normativity he set out to explain.

Post, Millikan, and Gibbard each turn outward to the biological world of materialism to find a source for normativity. What they find, in the end, are biological causes, and these are external to persons, indeed, imperiously so. That is, these causes are not subject to a person’s will, but are given to the person and unavoidable, applying independently of the person’s subjective motivational set—or, perhaps more accurately, are operative through a subjective motivational set that is determined by a diachronic causal skein. The moral “reasons” get lost in a host of non-moral causes. There are external causes, but no external authoritatively prescriptive moral reasons. With Millikan and Post, as with Gibbard, moral philosophy has been led into the pit of moral reasons arbitrariness. Moral reasons are set by evolutionary forces that meet in every individual’s diachronic causal skein.

In contrast to moral theorists like Millikan, Post, and Gibbard who turn outward to the materialist world to ground moral normativity, the great hope of contemporary Kantians is that the “pushmis” and the “pullyus” have (somehow) come apart and that authoritatively prescriptive moral reasons can be found by looking inward to the space thereby opened up. One must say that it does seem that we have stepped back (somehow) from the diachronic causal skein of evolutionary biology. In fact, for Korsgaard, this seeming is enough. It need not actually have happened; or perhaps the seeming that is embodied in a standpoint is (enough of) its happening. We do indeed respond to reasons, and both the reasons and the
responding to them seem to be *sui generis*. But this move—as I have attempted to argue, against Korsgaard—is itself barred for the materialist moral philosopher. If there is such a “separation” of the imperative from the imperative in human animals (itself a thorny explanatory problem for materialists, as Heidegger’s labors make clear), the freedom from all external givens also effectively isolates one from prescriptive moral norms that address us with authoritative reasons. And, in any case, the moral theorist who is pursuing this tack for avoiding moral reasons arbitrariness is no longer looking outward to biology for the needed moral resources, but inward to formal features of a freedom that has (somehow) arisen from out of the diachronic causal skein that holds all other animals stiflingly close. I have argued against Korsgaard, that (at least her version of) this path does not work, and, in any case, it is not the path that Post attempts to follow.

Post attempts to move from nature to norms by “pressing from below.” If my arguments are correct, the path of “bionorms” that Millikan has opened up and that Post has tried to follow has ended in the “coal pit” of moral reasons arbitrariness. On my diagnosis, this is because Millikan and Post have looked outward to a materialist biological world that gives the moral theorist nothing but non-moral efficient causality to work with. All the appeals to the diachronic nature of this causality and to the derivative proper functions arising for us as beings for whom much of this causality arrives in cultural forms do nothing to transmute the causes into moral reasons.

Insofar as this failure is characteristic of the outward turn of moral philosophers committed to materialism (which I do not argue here, but I do believe *could* be argued), materialist moral philosophers cannot look for help in avoiding moral reasons arbitrariness
from this corner. Given the problems of moral reasons arbitrariness, such philosophers should be highly motivated to explore all other options for another way forward.
CHAPTER SIX

CONCLUSION: BEING LED OUT OF A COAL PIT¹

*People* decide what to do. Faced with a moral decision, we consider, weigh, and judge what merits being done. We come to see what is uniquely justified by the reasons pertinent to the decision we must make, and we act *on* them (or *against* them). We recognize what is called for in a situation by appreciating the support relations that uniquely justify a course of action. But if the moral life essentially involves *people* acting *on* reasons that somehow stand over against them and authoritatively call for a particular response, moral reasons arbitrariness cuts the nerve of that life by stripping people of any authoritative address or summons to act one way rather than another. Nothing speaks *authoritatively* to them *as people*.

One path to moral reasons arbitrariness is to look outward and find only non-personal external *causes*, forces operating on people no matter what their subjective motivational set. But these bypass the person, much like what happens to Jim’s leg when the doctor taps his knee with a mallet. In this case, what had been taken to be reasons are exposed as operations of various non-moral causal forces, and the person who accepts that this is what is “really” going on finds that *she* is no longer addressed authoritatively. What shall she make of these causal forces? She has seen through the reasons to the non-moral causes. Seen in this light, moral reasons lose their authority to trump other forces operative

---

¹ I will speak a bit more freely here in the conclusion; the reader concerned about what I say here is referred back to the spadework of the earlier chapters. The careful work to support what I say here is in those earlier chapters.
in the situation. Where she had seen a particular action as authoritatively called for, she now sees only causal forces playing themselves out as they will.

Having seen the problems with this (as Kantians typically do), other philosophers look to the distance from the causal forces that is characteristic of the person as a person. As persons, we have “stepped back” from (as Nagel puts it) or taken up a “standpoint” (as Korsgaard would have it) of rational reflection about those causal forces. “Yes,” they say, “it is true that we cannot find reasons to act in the causal world. How could we? Unavoidably, we must decide what to make of the causal forces, and in this moment of deliberative necessity we find that we have stepped into the standpoint of moral deliberation.” The very fact of the reflection and the consequent necessity to act as a person in deliberative freedom from those causes itself instantiates the moral world in a moment something like what Kant called “the fact of reason.”

Where the outward turn fails, we inevitably turn inward and we find there the morality that could not possibly have come from without. In short, we address independent, authoritatively prescriptive reasons to ourselves. In Korsgaard’s terms, we confer value.

But this brings up problems of its own, most basically the question of what will constrain the act of conferring value. One person “confers” value on this, another on that. How will the authority and independence characteristic of moral reasons be supported? Korsgaard, as we have seen, delivers what Gibbard called an “elaborate Kantian package” in response to this question, and I have argued extensively that even these intricate machinations fail to do the work Korsgaard needs done. A uniform “human” identity appears chimerical when set against the backdrop of competing identities that gain credence.

against the backdrop of materialism. Real questions also surround the issue of how values arising in the individual necessarily and consistently bring other individuals within the ambit of those values since the first-personal performative pressures unavoidable in regard to oneself cannot similarly insinuate others into one's own reflections. Nor is it clear how the "elaborate package" needed to give the turn inward to the self even the slightest air of plausibility could realistically have arisen from the operation of evolutionary forces. In the end, independent, authoritatively prescriptive moral reasons dissipate across a welter of competing subjective motivational sets where people see that moral judgments are not uniquely justified. Having turned inward, we find ourselves in a pit of moral reasons arbitrariness once again.

Nowhere have I argued that the inward turn and the outward turn (and various hybrids) are the only possible places for moral philosophy to look for an answer to moral reasons arbitrariness. Neither have I made any claim to have treated all possible materialist versions of the inward turn and the outward turn or to have given a general argument why all materialist versions of these two ways of proceeding must lead to moral reasons arbitrariness—although we have seen Gilbert Harman claim, in effect, that they do. I have argued, however, that three materialist moral philosophies all succumb to moral reasons arbitrariness.

In different ways, Allan Gibbard and John Post both turn outward to the world of evolutionary forces, and both of them illustrate some of the difficulties that can be expected in this way of proceeding. Neither of them succeed in transmuting the non-personal causal forces they find out in the materialist world into the right kind of thing to address persons making moral decisions, or—in Post's case—it is done only by smuggling in the needed
normativity. Post begins by appropriating central elements of Millikan’s attempt to ground quasi-norms and tries to “press from below” on the problem of genuine normativity within a materialist world. One problem facing Post’s attempt to move from nature to norms is that the problems of Millikan’s own account of the normativity in language and intentionality devolve onto his own project, and there are a number of real difficulties besetting Millikan’s account. But even if her account were to succeed in its own terms, it would seem to be a most unlikely candidate for a foundation of a metaphysics of morals. Millikan appeals to the diachronic causal skein of human evolutionary history to ground proper functions, but such proper functions are not likely to accord well with moral duties—as the sociobiologists’ suggestion that maybe men have a “proper function” of abusing their stepchildren immediately makes clear. Post thus needs proper functions that somehow rightly trump other proper functions, rather than merely having greater force. He formulates a number of bridge principles that are meant to lead to this end, culminating in a principle he calls MORAL. But this not only smuggles in normativity in the form of assuming a “moral agent” outfitted with exactly the authoritative prescriptivity needed (only in an epistemic register), it also is down on all fours with other possible bridge principles, like VIABLE, which can claim the same provenance as MORAL. What Post does not show is how MORAL can trump VIABLE without smuggling in the very normativity that is being shown to arise from nature. Post ends up in the coal pit of moral reasons arbitrariness, and other philosophers tempted to press Millikan (or relevantly similar views) into service of moral philosophy seem likely to face similar difficulties.

Gibbard is more consistent than Post, but he ends up struggling mightily with how to rein in the “endorsement” he installs at the heart of norm-expressivism. Explicitly
eschewing descriptivist accounts as leaving out precisely the endorsement that is essential to morality, Gibbard gives a non-cognitive account of morality. Accordingly, moral reasons do not precede and guide endorsement, but rather *express* that endorsement. But then what will constrain endorsement? Gibbard labors to give an account of the objectivity of morality, but leaves us in a world where the “authority” of reasons is conversational pressure, and their “independence” is the fact that when people endorse whatever system of norms they endorse, they cannot help but think that others should “do so as well.” But any sincerely endorsed system of norms will enjoy such “independence,” and conversational pressure can be leveraged by anyone with suitable social capital or power. Gibbard recognizes that—given his materialist noncognitivism—“content-neutral” higher order norms and ideal awareness of facts fail to rectify the problem, and he turns finally to a “pragmatism of last resort” when the chips are down. The resulting picture of human moral life is queer—a misshapen representation of human recognition of reasons that rightfully call for us to take a particular course of action.

Where Gibbard and Post turn outward to ground their respective materialist moral philosophies, Christine Korsgaard turns inward toward a self accoutered with all sorts of abilities. But Korsgaard gives us no inkling of how all this “elaborate Kantian package” could plausibly have evolved in a materialist world. What is more, her moral philosophy vacillates between the two poles of the “identity/reasons” problem, depending on how she understands maxims. On the one hand, in the *identity-priority maxims* view, Korsgaard prioritizes practical identities over reasons. But then the identities themselves arise arbitrarily from non-moral forces and the threat of the Mafioso looms. This threat pushes

---

3 This was argued extensively in Chapter 3. The reader is referred there for the relevant details.
Korsgaard to appeal to an overarching “humanity,” but she struggles not only to square this with her materialism, but also to show why this identity should not be considered a shill for non-moral forces, how this identity compels de facto endorsement from all people, and why we cannot avoid taking others into consideration along with ourselves. And so, on the other hand, one might appeal to a form-priority maxims view that prioritizes reasons over identities. Now this would be a strange position for a materialist; after all, where would these reasons exist in the material world if not in evolved beings? Indeed, in Korsgaard’s materialist view, normativity is supposed to arise when a person endorses something. Moreover, the maxim-structure Korsgaard puts forward as an “intrinsically normative entity,” does not constrain willing by its form alone but requires the “human” identity that wills consistently. As Korsgaard realizes explicitly in The Sources of Normativity, there can be bad maxims, and something must winnow these out. But this is the identity that is supposed to be called out by the form of the maxim. So Korsgaard is cast back on the other pole of the identity/reasons problem. Neither pole solves the problem of moral reasons arbitrariness.

This circle, I suggested, manifests a typical Kantian structure of heautonomy. Korsgaard’s value-conferring agent seems caught in a circle of self-regarding structures that reach out to an objectivity that it knows full well it has not obtained. With the agent’s receptivity confined to the ambit of its own possibilities of understanding and thus always cut off from a purposiveness that addresses the agent from without, the heautonomous self finds itself inexorably drawn to a purposiveness perpetually forestalled. If this is accurate, the person sees that it is only as if she had moral reasons and that the fact that she inevitably takes them as objective owes to a subjectivity that fails to reach beyond the compass of the
individual herself. In short, I suggested that Korsgaard’s struggles to ground moral objectivity might reveal a deeper, structural weakness of attempts to avoid moral reasons arbitrariness by turning inward.

If I have successfully shown that there is a problem of moral reasons arbitrariness and that three materialist philosophies stumble into it, that is a significant result. Allan Gibbard and Christine Korsgaard have each done work in moral philosophy that is carefully crafted and widely regarded as setting out two important visions for moral philosophy. If they both succumb to moral reasons arbitrariness, then two leading visions of materialist moral philosophy have deeper problems than is widely recognized—problems that will also crop up for other moral philosophies similar to them in the relevant ways. John Post, though not as widely known, has explored some of the resources that Ruth Garrett Millikan’s materialist work in the philosophy of language and mind might offer for a materialist metaphysics of morals. Millikan is indisputably an important philosopher who has made a major and exceedingly subtle contribution to materialist attempts to work out what she calls “the normativity problem” as it shows up in language and intentionality. Post’s labors to bring some central elements of Millikan’s work to bear on moral philosophy merit careful attention as an attempt to leverage her work on a materialistically conceived “normativity” in the realm of moral philosophy—an attempt likely to be repeated by others in the near future. Here again, if my arguments are solid, the path Post has worked to open up and that others are likely to tread leads down into moral reasons arbitrariness. Such moral philosophies thus face a stiff challenge.

There are, of course, many other materialist moral philosophies that I have not addressed directly in these pages, and also many non-materialist moral philosophies as well.
Any such theory that would claim to do justice to the objectivity so obviously an essential part of the phenomenology of moral deliberation and action needs to face the problem of moral reasons arbitrariness squarely. Whether any moral philosophy—materialist or non-materialist (other than the three I have considered)—will be able to meet the challenge is an open question, and one not addressed here. For my part, I doubt that any materialist philosophy will be able to meet the challenge, though I have not argued that in this dissertation. What I hope I may have done along the way is set a pattern of some worthwhile and pointed questions that might be asked of materialist moral philosophies to put pressure on their attempts to salvage some sort of meaningful objectivity and avoid moral reasons arbitrariness. Let me attempt, briefly, to crystallize some of them and point out the challenge they may pose to materialists.

I. Some Questions to Ask of Materialist Moral Philosophies

After laying out the general plan of the dissertation in the Introduction, the first pages of the case I began to lay out against the materialist moral philosophies of Gibbard, Korsgaard, and Post carefully and vividly called attention to what it is like to live as beings who recognize and respond to reasons—what it is like to be a person. I take it that we all know a difference between what is happening when Jim’s leg jerks upward when the doctor taps his knee and what we do when working a geometry problem or deciding what is morally right. And a bit of attention reveals differences between the latter two as well, especially regarding the notion of rightful prescriptivity, with authority or commandingness attending only the latter. Again, geometrical akrasia would be weird, but I take it we are intimately familiar with it in the moral realm. Putting all this before our minds as clearly as
possible strikes me as a crucial starting point for putting pressure on materialist moral philosophies. Reducing or eliminating what it is like to be a person alive to moral reason is a stiff drink and not readily made plausible; but explaining all this and retaining the essential features of what we routinely do as people is difficult indeed within materialism. This is a tall order not only for the materialist moral philosophies considered here, but for materialist moral philosophies generally. They should be pressured to make sense of the most familiar features of human moral experience.

The problem is really that of accounting for people and what we do as a matter of course. A moral philosophy needs to show how persons are addressed with what Stephen Darwall helpfully calls “reasons of the right kind.” Reasons of the right kind need to be independent. Think of Jane working her geometry problem. The reasons for her to use the Pythagorean Theorem stand for anyone working the problem, regardless of whether they see them or not, and irrespective of the content of their subjective motivational set. They are in some real sense external to Jane—and everyone else. And so with moral judgments. Most anyone will find some point where they draw a line and say “This is wrong. Period.” Maybe it’s female circumcision or abortion or slavery or genocide. Indeed, there are some moral judgments that are right. And if some people won’t say it, they live it at many points nonetheless. A moral theory needs to show how it supports some substantive notion of the independence of moral reasons. And an important place to put pressure on materialist moral philosophies is whether they have such an account. Specifically, such philosophies are particularly prone to letting this independence dissipate across a welter of conflicting subjective motivational sets. Pressure should be brought to bear on materialist moral philosophies at just this point: How, specifically, does this philosophy find something more
fundamental than the wildly varying subjective motivational sets of various individuals, thus underwriting the independence of moral judgments? We have seen Gibbard, Korsgaard, and Post each foundering on this question.

But the independence of reasons is not enough. Reasons of the right kind to address people also need to be prescriptive and authoritative. The person who sees a moral reason recognizes that a course of action is rightfully called for. In addressing a person, a moral reason trumps or overrides other concerns. Ted may have more money and save himself trouble if he does not leave a note on the windshield; but that is irrelevant to what is morally right. He finds himself addressed with reasons that rightly override such concerns. He is called to leave a note, even if he has very strong motivational reasons not to. But such authoritative prescriptivity is not easily accounted for within materialist moral philosophies. External, non-moral causes operate on a person, bypassing him as a moral agent (think of Jim’s leg again), rather than becoming effectual through a person deliberating on them and recognizing that something is uniquely supported in a moral situation and that these support relations call for him to act. To fail to act in accord with this call or demand has the sense of rebelling against a rightful claim. The question, which I have pressed continually in this dissertation in various ways, is how the non-moral forces that predated and gave rise to human beings in a material world could be transmuted into reasons that address human beings and have rightful authority in doing so. How could non-moral causes ever even address people, much less be recognized by those people as having the authority to rightfully demand a course of action? Different materialist moral philosophies will try numerous approaches to this question. I have examined several and argued that they fall well short. The question is worth putting to other materialist moral philosophies as well.
When a materialist moral philosophy looks primarily outward, to the material stuff that precedes persons and will continue after we are gone, it is not difficult to see that the authoritative prescriptivity has gone missing. “Conversational demands” arising in response to evolutionary needs for cooperation with the “in-group” one happened to have been thrown into are a long way from having authority in the sense of a rightful claim. What if the group happens to be the “brown shirts” in Germany? Or consider Post’s attempt to find one “proper function” that can rightfully trump another “proper function.” They have the same provenance (MORAL and VIABLE would share the same biological pedigree). What makes one right instead of the other? One might have greater force, but this is simply not the same as having the rightful claim to the adherence of people as they deliberate. What if VIABLE typically has greater force and has sway over moral deliberations? Is that then right? Of course not. But precisely that is what Post’s moral philosophy struggles to support. In this he keeps company with Gibbard—and, I suspect, many other materialist moral philosophies.

This is where many philosophers turn inward to the specifically human to ground authoritative prescriptivity. The impulse driving them there should be applauded, and it can seem like there might be better hope of success. Surely we can have some sort of prescriptivity there: I tell myself what to do (or we tell ourselves). If authority can be brought into this picture somehow, it seems like we would have what we need. But a number of questions put themselves forward almost immediately. One, already mentioned above, is how this will avoid being an exercise in the unfettered expression of a welter of differing subjective motivational sets (including one’s own). Kantian philosophers in particular will be ready with a number of replies, but we have seen that these are open to
some very damaging objections in the case of Korsgaard. It seems likely that similar problems might dog the moral philosophies of a Nagel or a Rawls or a Habermas. At least the questions would be worth asking. Additionally, I argued that Korsgaard’s version of the appeal to “performative necessity” to drive a kind of “universalizability” argument ends in failure because those who are not me cannot be brought into deliberations in the same first-personal way I cannot help but take up with respect to myself. Simply put, I do not need to take them into consideration in the same performative way I cannot help but take myself into consideration. Now if I accepted a metaphysically thicker understanding of human beings than the austere metaphysics of materialism will allow, such an argument might have purchase. But the whole point of talking about performative necessity is precisely to avoid the need to appeal to something more robust. Something about the pattern or procedures of my self-regarding thought is supposed to be enough. At least in Korsgaard’s version of this move, however, it is not. Perhaps it fails in other versions of the inward turn to human subjectivity as well.

An inward turn like Korsgaard’s also faces problems in squaring itself with the strictures of materialism. To put the question bluntly: Can something like Korsgaard’s “elaborate Kantian package” really be thought to have evolved in a materialist world? Can maxims be intrinsically normative entities that exist, somehow, independently of evolved beings that think them? Materialist versions of the inward turn like Korsgaard’s must abide by the strictures of materialism. It is worth pressing them on this point. Thus, when Thomas Scanlon takes reasons as “primitive,” this shows good sense—but he gives no inkling of how they could be, given his materialism. This is quite unsatisfactory. Moreover, views relevantly like Korsgaard’s owe a story of why it is that the favored “human” identity
(whatever it is called) is not itself a shill for non-moral causes that has, as such, no unique claim to be pursued over other possible identities. A person who accepts materialism may have good reasons to think that the “humanity” that Korsgaard advances as set apart from all other identities is just one more vehicle for the operation of non-moral forces and that its attraction for us can be explained in those terms. As such, other viable identities stand on the same footing. The reasons endorsed within these identities will be thought to be arbitrary.

This question of the arbitrariness that arises when conflicting moral judgments can claim the same justificatory grounds with equal right has been advanced in a number of different ways in this dissertation. The question is whether a materialist moral philosophy uniquely justifies some moral judgments, or if it leaves a number of conflicting moral judgments on the same footing. Seiriol Morgan’s excellent article, “Naturalism and Normativity,” provides a first-rate example of this problem in the thought of “synthetic realists” like Sturgeon, Railton, and Brink. The same procedure that supposedly establishes “moral realism” as an objective realm of discourse also establishes a Nietzschean “noble realism” on the same grounds. The problem, I have argued, is that human nature, materialistically conceived, cannot support the justificatory weight, so that what is justified on grounds of explanatory necessity will vary with who is doing the explaining. Be that as it may, we see in Morgan’s essay a critique that presses yet another group of materialist philosophers with the problem of moral reasons arbitrariness. The question that he presses in that article is one that I have also pressed here with Gibbard, Korsgaard, and Post. My guess is that it could profitably be asked more widely still.
I have pressed these points with a number of examples: the viable dictator, the contented criminal and the Mafioso, the ideally coherent anorexic and the slaveholder. Employing concrete examples like these also helps put some needed pressure on materialist moral philosophies. Do they address the moral questions arising in these views convincingly, or as they proceed do we gradually see that the view of the moral life they present is strangely disfigured? Examples that press questions of coherence, full factual information, and viability helpfully give a focal point around which questions like those above can be developed and made clear in their implications.

I have not argued that questions similar to those I have asked in this dissertation will show that all materialist moral philosophies succumb to moral reasons arbitrariness, but I do think that their critical import as displayed in relation to the work of Gibbard, Korsgaard, and Post should lead to further investigations. Do similar questions show that other materialist moral philosophies also fail to reply convincingly to moral reasons arbitrariness? If they do, are there general reasons for this failure? Would those reasons perhaps extend to other aspects of normativity, such as found in aesthetics, epistemology, mathematics, language, and intentionality? Would some non-materialist moral philosophies be better situated to reply to the moral reasons arbitrariness, and, if so, what could be said for these moral philosophies? On the other hand, could materialist moral philosophies make do with a less robust sense of moral objectivity and give us all the objectivity we practically need and (in any case) could reasonably hope for? These are questions that might merit additional exploration in light of the arguments presented in this dissertation.

---

4 See Cuneo, The Normative Web, for an argument that the normativity in moral judgments and epistemic judgments cannot be separated in the way many contemporary philosophers have assumed.
II. Two Concluding Questions

A. How “Modest” Can Moral Objectivity Be before It Is Meager?

Before concluding, I would like to address the last two of these questions very briefly, starting with the question of whether materialist moral philosophies could get by with a less “grandiose” objectivity (as Gibbard would put it) than what I have argued that Gibbard, Korsgaard, and Post have failed to capture. Part of a reply might be to return to the arguments against the moral philosophies examined here, pointing out that extensive argumentation has been presented to show that the kind of moral “objectivity” in these philosophies sinks below “modest” to meager. These arguments have already been presented in detail and summarized in this conclusion, and there is no need to repeat them again. They should, however, be borne in mind as one asks questions about how thin an objectivity one can make do with.

That said, the question will obviously still be pressed with vigor by some, for the stakes are high—perhaps even as high as a choice between materialism and moral objectivity. We should expect to be told that we can make do with a fairly thin moral objectivity, one that can do without independent, authoritatively prescriptive moral reasons. In response to this, I will not pretend to give a one-size-fits-all answer. Indeed, it strikes me that the devil will be in the details with the proposed substitutes, and the arguments above have exposed just how devilish some of the details can be. And yet there may be a pattern of questions that might be counted on to expose weaknesses in the various “modest” objectivities that are proposed. Let me suggest that this pattern will be tied to the nature of

5 These questions have already been considered somewhat at pp. 234-35 in connection with Gibbard’s program of “modest” objectivity.
moral reasons as those were seen to be embedded in the ways people think and talk about reasons.

At the start of Chapter 2, careful attention to the ways people think and talk about reasons revealed the important role of independent, authoritative prescriptivity in moral reasons. A candidate “modest” objectivity will massage the independence, the authority, or the prescriptivity of reasons (or some combination of them), and my suggestion is that we put pressure on them at these points to see how they stand up. Specifically, what does this “modest” objectivity look like? In what way is it independent of the subjective motivational sets of various contingent personal beings? Does it address people as people, so that they can recognize that a particular judgment or action is called for in a moral situation? In what sense does it have authority in these matters, rightfully overriding other concerns? How does it fare in relation to concrete cases like slavery? In each of these questions, we can ask what happens in the massaged version of moral “objectivity.” Do we have a recognizable version moral deliberation, or does it appear misshapen in some ways?

Consider independence, for example. If a person expresses the moral judgments she does because they are accepted by a group of people that she happened to be thrown into and whose cooperation she needs if her projects are to succeed, her moral judgments would not be marked by a very interesting sense of independence. Thus if someone states his condemnation of militant Pakistani Muslims and says that this is his attitude, the attitude of his group that he hopes is shared by many, the independence of his moral judgment is meager.6 People with conflicting moral judgments can claim the same independence on the same grounds. Indeed, the parties to one of Gilbert Harman’s “implicit agreements” could

---

6 See pp 132-33 above.
claim independence in that sense. Such “independence” might be enough to ensure cooperation among the members of an in-group that will help their projects to succeed, but this is not the kind of independence in view in moral reasons. It is not the kind of independence Jane understands her reasons for using the Pythagorean Theorem to have; nor is it the kind of independence Ted understands the reasons he should leave a note on the windshield to have. We are faced with moral demands that we do not understand as hinging on our own subjective motivational set or on the non-moral contingencies of the group of people whose cooperation we need. As Paul Moser notes about moral demands, “some of these demands go against our preferences, including our selfish preferences. They don’t arise uniformly from our individual wills or even from the common will of our peer group.” Thus the moral reformer stands over against the group because he sees what is called for, even if it costs him the success of his own projects. A moral philosophy that undercuts this robust sense of independence fails to do justice to a critically important part of moral deliberation, judgment, and action.

Such failure is not the failure of a “grandiose” objectivity that “may not be vastly important.” Nor does the failure have to do with “strange beings who are merely conceivable.” The failure undermines a basic element of what people take themselves to be doing when they make a moral judgment. As Allan Gibbard moves away from “grandiose” objectivity, he recommends restricting our formulation of moral judgments to “actual

---

7 Moser, The Elusive God, 54.

8 Gibbard, Wise Choices, Apt Feelings, 201.

9 Gibbard, Wise Choices, Apt Feelings, 201.
rational beings.”\textsuperscript{10} Well and good. But then he asks, “once we restrict our attention to actual beings, why not consider actual beings in the vicinity, the ones with whom we might discuss and interact?”\textsuperscript{11} An answer springs to mind: Because what is their authority? That they are local? That they are “indispensible” for my projects (I’d like them to help me build my home, or at least not burn it down)? Something more will need to be said if what Gibbard says here is even to get in the neighborhood of what we mean by independence. If my moral judgments depend on such contingencies, the independence of moral judgments is a complete wash. Gibbard will need some additional story (perhaps a more robust account of higher order norms and full information that leaves his materialist noncognitivism behind). What I am recommending is that we put pressure on whatever those additional stories may be, in something like the way I have done in this dissertation.

Undoubtedly we could say more regarding questions of just how much a materialist moral philosophy can massage the independence, authority, and prescriptivity crucial to moral reasons without becoming meager. But perhaps what has been said here will make a start on these issues and be suggestive of how further inquiry might proceed.

B. How Might a Non-Materialist Moral Philosophy Address Moral Reasons Arbitrariness?

Another issue worth further exploration would be how non-materialist moral philosophies might fare with respect to moral reasons arbitrariness. Pretty clearly, such moral philosophies have different resources to meet the problem, though they too must answer the question of moral reasons arbitrariness and will need to face other challenges.

\textsuperscript{10} Gibbard, \textit{Wise Choices, Apt Feelings}, 203.

\textsuperscript{11} Gibbard, \textit{Wise Choices, Apt Feelings}, 203.
Theistic moral philosophies can appeal to aspects of their ontology that afford promising avenues for meeting the challenge of moral reasons arbitrariness. In particular, this is certainly the case with the family of theistic moral philosophies I am most familiar with, namely, moral philosophies that take into account the Triune God of Christian faith and the relevance of Jesus for philosophy—moral philosophy in particular. Let me reflect briefly and suggestively on how some of what is available within the Christian moral philosophies I know best might make a difference in defusing the challenge of moral reasons arbitrariness, beginning with a whirlwind overview of some central Christian claims about what reality is like. This overview will sketch some resources one particular non-materialist moral philosophy might have available for a reply to the problem of moral reasons arbitrariness.

First and foremost, if Christianity is true, there has never been a time when one God did not exist as three persons in perfect relationships of love. Before all else came not only personhood, but the perichoretic dance of three perfect persons, each loving and taking joy in the other without alloy. “God is love” (I John 4:8, 16; New International Version). From within the perfection of personhood, three persons have always addressed one another. When from His abundance God created, all things were created “by him and for him” (Colossians 1:16, referring specifically to the Son of God) and thus all created things bear the stamp of his purposes—purposes of loving relationship. Some beings made in His image were created with the capacity not only to point to Him but to enter into relationship with Him and with each other. They were made in His image, reflecting the pattern of His being as a person. The capacity to love was and is their highest capacity, and precisely in the misdirection of this capacity they became disfigured and ruined. In Augustine’s poignant phrase, they “loved perversely,” so that in turning from God and twisting their own being,
they “turned to themselves whose being is relative—a sin that can have no better name than pride.”12 Thus they set themselves against reality itself and abandoned that which was wholeness and goodness for beings who were made personal and relational. Made for loving relationship they exchanged God for things they loved above God (that is, idols) and so spurned their “Immutable Good which is so completely their good that, without this good, misery is inevitable.”13 Among people there was “no one who seeks God. All have turned away,” (Romans 3:11-12) and what remained to them was the wrath of a cuckolded God. Having turned from God in idolatry/adultery, “all [their] righteousness is like filthy rags” (Isaiah 64:6), like a man caught in adultery who cleans the dishes and expects his wife to be impressed.

But God in His love sought those who went whoring on Him (Hosea 2:13-3:5). And so the Father called the Son, and the Son responded to the loving and grace-filled call of the Father and, having been sent by Him, took on flesh and became a man who could be seen and heard and touched (John 1:1-18; I John 1:1-2). He came in God’s love, and where the old covenant relationship had been broken by our whoring, Jesus as a human being kept the human side of the covenant and paid the price of human whoring. In Him God set in place a new covenant relationship. As Jesus’ followers recalled his words during a memorable last meal before He was crucified: “This is my blood of the covenant which is poured out for many for the forgiveness of sins” (Matthew 26:28; see also Hebrews 7:11-8:13). Or, as one of those same disciples put it later in life: “This is love: not that we loved

---


13 Augustine, *The City of God*, XII, 1, p. 245. The language of exchanging God for idols is from Romans 1:23.
God, but that God loved us and sent his Son as an atoning sacrifice for our sins” (I John 4:10). Because of His love and his passion to see God’s name honored, Jesus was crucified, choosing to lay down His life for people for the forgiveness of sins (John 13:31-32, 17:4-5, 10:14-18). And so to the one “who does not work but trusts God who justifies the wicked, his faith is credited as righteousness” (Romans 4:5). But those who were with Jesus at that final meal (and many others), testified that Jesus rose from the dead and were willing to pay with their lives to let others know what they had seen. Their lives radically changed, and they quickly began taking their message and costly actions of reconciliation across entrenched racial and social barriers (see Acts). They called wayward people to “repent and turn to God and prove their repentance by their deeds” (Acts 26:20).

The last words of the Bible depict a reality of everlasting covenant faithfulness, love, and enjoyment of which the relationship between the husband and wife is a faint picture (or sacramental participation in): the Church is “the bride, the wife of the Lamb” and the “dwelling of God is with people, and he will live with them. They will be his people, and God himself will be with them and be their God. He will wipe every tear from their eyes. There will be no more death or mourning or crying or pain, for the old order of things has passed away” (Revelation 21: 9, 3-4; cf. Ephesians 5:21-33). And they will not taste “the second death” (Revelation 21:8), but will enjoy loving relationship with God and others without end.

Now, what does any of that have to do with replying to moral reasons arbitrariness? In short, if it’s true, then there seem to be some hopeful avenues in dealing with the problem. Grant materialists their metaphysic, and there is quite a bit of work to be done to avoid moral reasons arbitrariness—if it can be done at all. Grant Christians their
metaphysic, and moral reasons arbitrariness may be open to some fairly straightforward answers. This, of course, is more than I can argue extensively here, nor is it the burden of this dissertation to do so. But perhaps a few comments will suggest some of the power such a reply might have.

Let’s shape this sketch around the independence, authority, and prescriptivity that have guided these reflections since our first reflections on moral reasons. First, consider prescriptivity—reasons of the right kind to address people as people and to call for something to be done. As Stephen Darwall has seen (even noting its Christian provenance), the second-personal dimension is crucial here. People address people with reasons. In a materialist world, people scud momentarily across the surface of the fundamental, non-personal forces. These forces do not address people. “Nature, being unintelligent, doesn’t give me commands to love others, or any command for that matter, even if it gives me demonstrable bumps and bruises against my will. Despite setting limits on us and our efforts, nature remains altogether silent on moral injunctions, and other injunctions too… I am under no command (beyond possible merely human commands) to use my time with the priority to love others.”\(^\text{14}\) Clearly this is the case, granted materialism. But if the ontology of the Bible is true, then personhood does not bob tenuously on the surface of a non-personal ontological reality, but precedes and gives rise to all else. “In the beginning, God…” (Genesis 1:1). Moreover, if God is Triune, the second-personal address fundamental to moral reasons and—more basically—to love, has always been.\(^\text{15}\) When we


were called into existence, the Logos already addressed us, both in our very created form and in his address to us as beings who could recognize and respond to His call. Reasons of the right kind have always been, and we were made to recognize and respond to them. We were made to love, as God is love. Given this, prescriptivity is not a problem.

Now of course a major problem with people addressing people with reasons, is that these reasons tend to be all over the board. Given the vagaries of various human subjective motivational sets, they lack both independence and authority. Darwall recognizes this, and he finally tries to bring them back under the thumb of the Kantian self that autonomously prescribes the moral law to itself. In a world of materialist evolution of human beings, however, it’s far from clear that such beings will uniformly prescribe that to themselves, or that those who do not prescribe this in such a way as to trump other practical identities (think Mafioso) would have something wrong. But if the God portrayed in the Bible exists, then one, perfectly constant, perfectly whole person stands over against all the caprice of human prescriptive address. “Jesus Christ is the same yesterday, today, and forever” (Hebrews 13:8). Reasons addressed to us by Him stand independent of the subjective motivational sets of all contingent personal beings, addressing all other people with Love’s Demand.\(^{16}\) Indeed, reality itself is this God who is holy love; we could sooner avoid gravity than this call. No matter what anyone is motivated to do, they are addressed with reasons to make moral judgments consistent with love and God’s undying insistence on its integrity through all creation. Reality is most basically personal and relational in perfect integrity, and so the call to be in right relationship to God and others in perfect love stands over against all contingent personal beings in complete independence of all of them. To betray love is to

fling oneself against an unyielding reality, not only ignoring God’s address to all people, but also fouling the beauty of our own design as those who are, like God, persons who can love—and thus hear and respond to the reasons other persons give us.

Our design as persons also speaks to the authority we recognize as persons. The source of God’s authoritative address to all people is not adventitious to Himself, but just is His own being as perfectly personal and relational. Again, this just is reality. In the infamous Euthyphro problem, one of the “horns” of the dilemma is that God might will the pious because it is pious, the idea being that this standard is thus adventitious to God and we should look to that standard. But what is the ontology of that standard? In what does it consist? Perhaps the standard is personhood and relationality itself, and the ontology just is the existence of these in their perfection. Perhaps Plato tried to drive a merely epistemic wedge in something that can only be an ontological whole if it is at all. As beings who are personal, we fall under this standard by our very being—ontologically, we flourish only as we are wholes of our kind. That is, we flourish only as we leave our rebellion and return to the integrity of personhood. Or, to use imagery deeply woven into the fabric of the Bible, we flourish only as we turn from our idolatry/adultery to a God who is willing to enter into a new covenant of love with us. An idol is something we aim at over and above relationships of love—typically ourselves (the sin of pride). This attacks not only reality, but our own nature. By this nature we fall under an authority that is relentlessly and with fierce integrity, personal and relational. In this God is different than us, standing over against us as the perfection out of which we, as persons, find our pattern and our being. Perhaps the epistemic possibility of recognizing an authority that addresses us as persons rests on the call
of a prior ontological integrity of personhood and relationship. “We love because he first loved us” (I John 4:19).\textsuperscript{17}

Obviously, much more would need to be said to develop a workable moral philosophy from something like the resources sketched here, and there are many additional ways the resources I sketched might be relevant (think of Jesus crucified and risen), but perhaps enough has been said to suggest that there seem to be reasons to have some hope that such a moral philosophy would deal gracefully with the problem of moral reasons arbitrariness. But however much there is to be said for such a moral philosophy, it will be a non-starter for many philosophers today, among whom there is often a strong commitment to materialism. On purely intellectual grounds, it is difficult to see why it should be any more plausible that what came before all else was non-personal material stuff rather than a personal reality. There is nothing inherently more remarkable about a personal reality existing from all eternity than there is about a non-personal reality existing from all eternity. And if there were such a personal reality, it would not be surprising if it had an unswerving commitment to the integrity of personhood and relationships (that is, love), much like what rests at the foundation of Christianity. It could be that the matter is not so much intellectual, as a matter of the will—questions precisely of morality, authority, command, and submission.\textsuperscript{18} Perhaps in the very demand that God submit evidence for our

\textsuperscript{17} I do not, of course, pretend to have resolved the Euthyphro dilemma in one paragraph. Obviously, additional questions remain. That said, much good work has been done on this question, including Robert M. Adams, “A Modified Divine Command Theory of Ethical Wrongness,” in \textit{Morality and the Good Life}, ed. Thomas L. Carson and Paul K. Moser (Oxford: Oxford University Press, 1997), and Adams, \textit{Finite and Infinite Goods}, especially chapters 11 and 12. See also, Carson, \textit{Value and the Good Life}, 245–48.

\textsuperscript{18} Søren Kierkegaard put it this way:

People try to persuade us that the objections against Christianity spring from doubt. That is a complete misunderstanding. The objections against Christianity spring from insubordination, the
inspection, we attempt to ward off the way God does present Himself to us. But if God is indeed perfect in personhood and relationship, we would expect God to have purposes in making Himself known that would uphold the integrity of personhood and relationship. Furthermore, if this God sought to reconcile people who attack personhood and relationship—both in ourselves and others—we might expect that His purposes in making Himself known would not cater to the “cognitive voyeurism” of those creatures, but would seek their transformation. 

“The God in question would come to us with authoritative evidence of the divine reality, that is, evidence demanding that we yield our wills to (the will of) the divine source of the evidence in question.” In making Himself known, He would uphold the integrity of personhood and genuine (that is loving) relationship—He would, in other words, knit reality together and seek to make us whole again as people, reuniting the idolaters/adulterers with Himself.

The authoritative evidence in question would thus inevitably have some moral dimension to it, a moral dimension that might not always be welcomed. And yet we meet
here with a kind of weightiness that cannot be gainsaid, and perhaps more than anything it is the weight of authoritative evidence that makes a moral claim on us that should be thrown into the balance against moral reasons arbitrariness. We see this when Thomas Nagel makes his most forceful argument against “attempts to get entirely outside of the object language of practical reasons, good and bad, right and wrong, and to see all such judgments as expressions of a contingent, nonobjective perspective.”\textsuperscript{22} Such attempts, Nagel submits, “will eventually collapse before the independent force of the first-order judgments themselves.”\textsuperscript{23} Or again, Emmanuel Levinas makes much the same point, but with more depth: “The first word of the face is the ‘Thou shalt not kill’. It is an order. There is a commandment in the appearance of the face, as if a master spoke to me.”\textsuperscript{24} Here is the voice of independent, authoritatively prescriptive moral reasons. These claims are real and they come with an authority that stands over against us and calls for “costly, observable love.”\textsuperscript{25} They are expressions of Love’s Demand, a demand rooted in the Divine Logos who was and is and is to come—a demand upheld in perfect integrity by reality itself.

I have argued that the contemporary materialist moral philosophies of Allan Gibbard, Christine Korsgaard, and John Post do not fit with the “commandment in the appearance of the face.” Instead, they threaten to abet the contemporary (and ancient) suppression of “Love’s Demand” as it comes to local expression in the face of even the smallest and weakest of people, dissipating either the independence, the authority, or the

\textsuperscript{22} Nagel, \textit{The Last Word}, 103.

\textsuperscript{23} Nagel, \textit{The Last Word}, 103; emphasis added.


\textsuperscript{25} Francis Schaeffer, \textit{The Mark of the Christian}, 2nd ed. (Downers Grove, IL: InterVarsity Press, 2006).
prescriptivity (or some combination of them) essential to the moral reasons they would rightly give us. My language for this is that moral philosophies like these succumb to moral reasons arbitrariness.

I conclude with a final discomfitting question about moral reasons arbitrariness.

Here is how Paul Moser raises the issue:

Our failure to apprehend God’s authoritative call saliently may result from our preferring not to apprehend it on God’s terms of unselfish love for all people. We often prefer, for instance, not to have to forgive or to love our enemies, and we act accordingly toward our enemies... It seems easier, or at least more in our own interest, to suppress or to ignore any call from God for us to live as dependent obedient children of God who reflect, if imperfectly, perfect divine love. God’s authoritative call toward perfect love would be anything but comfortable, given our selfish ways, particularly if we resolve to obey, come what may.26

Perhaps this is something that tempts us all, issuing in a “cosmic authority problem,” as Nagel has put it.27 Nagel aversthat this authority problem might be responsible for much of the “scientism and reductionism of our time.”28 Perhaps he is right, and we might well add: “and much of the functional deism and religious hypocrisy of all ages.” Perhaps we like to take shelter from God’s unwavering commitment to and insistence upon the integrity of personhood and relationship by hiding behind moral reasons arbitrariness. Religious shelters are some of the best, but there are also philosophical shelters. Philosophically, it may be that our continual attraction to moral reasons arbitrariness—at least when it gets us off the hook—drives a commitment to materialism, rather than the other way around. I have not argued that this is the case, but if it is so, it might be very difficult indeed to unseat the metaphysical commitments and be led out of the coal pit moral philosophy is in. In that

---

26 Moser, The Elusive God, 51; emphasis in original.

27 Nagel, The Last Word, 131.

28 Nagel, The Last Word, 131.
case, the only way out of the coal pit would be as we are led out by God’s authoritative self-revelation in conscience and wherever else He bends down to us.
BIBLIOGRAPHY


______. “Morality as Consistency in Living: Korsgaard's Kantian Lecture.” Ethics 110 (October 1999): 140-64.


________. “Kant on Recognizing Our Duties as God’s Commands.” Faith and Philosophy 17, no. 4 (October 2000): 459-78.


Korsgaard, Christine M. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press, 1996


VITA

Brad Seeman was born and raised in Gurnee, Illinois. He attended Wheaton College before graduating magna cum laude from Trinity Evangelical Divinity School with a Master of Arts in Church History and Philosophy of Religion. While at Trinity, Brad was awarded the T.B. Madsen Award for an Outstanding M.A. Thesis in Church History. Brad also holds the M.A. in Philosophy from Loyola University Chicago.

Brad has published five scholarly articles in refereed journals, including International Philosophical Quarterly, Philosophia Christi, and International Journal for the Philosophy of Religion. He has also presented papers at a number of conferences. His research interests include metaethics, philosophy of religion, the intersection of analytic and continental philosophy, architecture, and Church history. Brad is currently working on an essay related to his earlier work in Church history to be published in a forthcoming volume for Eerdmans.

Brad holds the position of Assistant Professor of Philosophy at Taylor University in Upland, Indiana. He lives in Jonesboro, IN with his wife, Kiersten, and their five children, Grace, Emma, Luke, Elias, and Josiah.
DISSEMINATION APPROVAL SHEET

The dissertation submitted by Bradley Nelson Seeman has been read and approved by the following committee:

Paul K. Moser, Ph.D., Director
Professor of Philosophy
Loyola University Chicago

Thomas L. Carson, Ph.D.
Professor of Philosophy
Loyola University Chicago

Andrew Cutfellos, Ph.D.
Professor of Philosophy
Loyola University Chicago

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the committee with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

9-4-09

Date

Paul K. Moser

Director’s Signature