FOUNDATIONALISM AND THE FOUNDATIONS OF ETHICS

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Irfan A. Khawaja

David Solomon, Director

Graduate Program in Philosophy
Notre Dame, Indiana
July 2008
FOUNDATIONALISM AND THE FOUNDATIONS OF ETHICS

Abstract

by

Irfan A. Khawaja

The problem of the “foundations of ethics” is one of the oldest and most venerable in philosophy, discussed in a literature that stretches from the Platonic dialogues to the present. It is also, despite its currency, a remarkably obscure and multifarious problem: it’s rarely clear in the literature what “the” problem is, what the solutions are, and indeed, what counts as an instance of either problem or solution. The basic aims of this study are to defend a particular conception of the problem; to offer the sketch of a solution; and to offer an extended critique of one attempted solution, namely, Christine Korsgaard’s in *The Sources of Normativity*. Topics discussed include foundationalism as a general theory of epistemic justification, intuitionism as a moral epistemology, and meta-ethical approaches to the foundations of ethics (notably Kantianism and egoism). Principal figures discussed include Aristotle, Kant, Mill, Moore, Rawls, and Korsgaard.
# CONTENTS

Acknowledgments.............................................................................................................. vi

Introduction: Foundationalism and the Foundations of Ethics: The Problem .................1
  0.1 Introduction .......................................................................................................1
  0.2 Foundationalism as a General Theory of Epistemic Justification .....................2
  0.3 Conceptualizing Foundationalism about Ethics.................................................5
    0.3.1 The Lexicographic Strategy ................................................................7
    0.3.2 The Theory-Driven Strategy ..................................................................10
    0.3.3 Multiply-Instantiated Regress Arguments in Ethics .............................11
    0.3.4 The Foundations of Ethics as an Epistemic Project ..........................16
    0.3.5 The Foundations of Ethics as a Meta-Ethical Project .......................19
  0.4 The Plan of the Present Study ..........................................................................22
  0.5 Notes ................................................................................................................25

Chapter 1: Foundationalism, Epistemic and Meta-Ethical: A Taxonomy of Conceptual Possibilities .........................................................................................31
  1.1 Introduction ......................................................................................................31
  1.2 Homonymy ......................................................................................................34
  1.3 Analogy ............................................................................................................35
  1.4 Reductionism ...................................................................................................39
    1.4.1 Epistemic Reductionism ...................................................................39
    1.4.2 Ethical Reductionism ............................................................................43
  1.5 Epistemic Reductionism and Analogy .............................................................45
  1.6 Notes ................................................................................................................49

Chapter 2: The Epistemic Approach to the Foundations of Ethics ....................................54
  2.1 Introduction ......................................................................................................54
  2.2 Epistemic Foundationalism: Back to Basics ....................................................56
  2.3 Task (1): The Meta-Epistemology of Epistemic Justification .........................61
  2.4 Tasks (2) and (3): Epistemic Foundationalism, Genus and Species ..........70
  2.5 Task (4): Applying Foundationalist Strictures to the Moral Case ...............76
  2.6 Aristotelian Empiricism as a Foundationalist Moral Epistemology ..........87
  2.7 Notes ................................................................................................................90

Chapter 3: The Meta-Ethical Approach to the Foundations of Ethics ................................97
  3.1 Introduction ......................................................................................................97
  3.2 Meta-Ethical Foundationalism: Some Preliminary Caveats .........................99
ACKNOWLEDGMENTS

This dissertation took an inordinately long time to write, and so I owe a deep debt of gratitude to the University of Notre Dame, and especially to the Department of Philosophy for waiting for me to finish it. Patricia Blanchette, Montey Holloway, David Solomon, Leopold Steubenberg, and Tracy Westlake deserve special thanks for their patience, and for their help in moving things along.

Many of the ideas that found their way into the dissertation were initially conceived at Notre Dame, and I’d like to thank David Solomon again, as well as Alasdair MacIntyre and James Sterba, for valuable feedback on versions of chapters 1-6.

Most of the dissertation was written while I lived in Princeton, New Jersey and taught at The College of New Jersey (TCNJ). I’d like to thank all of my colleagues at TCNJ, but especially Richard Kamber, Pierre LeMorvan, and Consuelo Preti for their feedback and support. I owe a particularly large debt to Allan Gotthelf, then at TCNJ but now at the University of Pittsburgh, for philosophical conversations that influenced the whole dissertation, but especially chapters 4-6. I owe a similar debt to James Lennox also of Pittsburgh, for helpful comments on matters pertaining to explanation and to the philosophy of biology. And I owe both philosophical and personal debts to Hilary Persky, whose support made the dissertation possible, and whose comments enriched it.

I tested early versions of the Introduction as well as chapters 1, 2, 7, and 8 on members of the McCormick-Taber Seminar in Philosophy. Carrie-Ann Biondi, Teresa
Celada, Jason Raibley, and Michael Young all had interesting and useful things to say about the hundreds of pages I asked them to read; Michael’s extensive written comments on that material saved me from many errors, and forced me to consider issues I would never have thought of on my own.

I can’t forbear from mentioning three eating establishments whose (often free) food and drink kept me alive for extended stretches of time, and gave new meaning to the phrase “bon appétit”: Lula’s Café in South Bend, and Tiger Noodles and Hoagie Haven in Princeton. Socrates should have been so lucky.

I owe a special thanks to my wife, Carrie-Ann Biondi, whose substantive and stylistic comments, editing, proofreading, formatting, support, and love made the task of finishing the dissertation so much easier—and made the final product so much better—than it would otherwise have been.

Naturally, the responsibility for any remaining errors is mine alone.
INTRODUCTION:

FOUNDATIONALISM AND THE FOUNDATIONS OF ETHICS:

THE PROBLEM

0.1 Introduction

The problem of the “foundations of ethics” is one of the oldest and most venerable in philosophy, discussed in a literature that stretches from the Platonic dialogues to the present. It is also, despite its currency, a remarkably obscure and multifarious problem: it's rarely clear in the literature what “the” problem is, what the solutions are, and indeed, what counts as an instance of either problem or solution. The basic aims of this study are to defend a particular conception of the problem (Chapters 1-4); to offer the sketch of a solution (Chapters 5-6); and to offer an extended critique of one attempted solution, namely, Christine Korsgaard’s in *The Sources of Normativity* (Chapters 7-8).

The key to clarifying the problem, I argue, is to clarify the connections among *foundationalism as a general theory of epistemic justification, foundationalism as a moral epistemology, and foundationalism as a view in meta-ethics*. The task as I see it is partly a matter of supplying a rationale for the subject, and partly a matter of integrating—i.e., identifying connections between—existing rationales. To what question
is “a foundation for ethics” a solution? If there is more than one question, what relation, if any, do they bear to one another? And what counts as a good answer?

0.2 Foundationalism as a General Theory of Epistemic Justification

In English-speaking analytic philosophy, the term “foundationalism” usually names a theory of epistemic justification according to which an individual’s beliefs (or more precisely what we may call his cognitive set, including his beliefs and non-doxastic cognitions), when justified, have a double-tiered structure: foundational items confer justification without needing it, while superstructural items are justified insofar as they are supported by the foundation. Understood in this way, foundationalism is a solution to the “canonical” regress problem articulated in Aristotle’s Posterior Analytics I.3, and a rival of skepticism, coherentism, and contextualism.¹

Though there’s a great deal of contention in this literature, there is also a measure of agreement as to the basic formulation of the problem that gives rise to foundationalism, as well as the sort of solution that qualifies as foundationalist. Take a belief $p$ held by a subject $S$. Assume that $p$ is not self-justifying, and that $S$ is justified in believing $p$. Then for any $p$ and any $S$, we can ask why $S$ believes that $p$ on a specifically epistemic interpretation of “why,” and expect an answer. Call the answer the chain, and assume that the chain responds to the question, has at least one member, is asymmetric, and (assuming at least two members) is transitive.

Either the chain ends or it doesn’t. If it doesn’t end, we’re faced with the prima facie difficulty of explaining how the chain can (a) have any actual infinity of members,
while (b) justifying a given belief in finite time held by a person with finite cognitive powers. If it does end, we have to explain where it ends, how, and why.

Given the preceding, a foundationalist epistemic theory has three basic tasks, and yields three differentiae for different species of epistemic foundationalism. First, it has to give a plausible, non-arbitrary, non-circular account of the terminus of epistemic justification. Foundationalist theories differ on what is at the foundations and why: On some accounts, the foundation is doxastic, on others, not; on some the foundation enjoys a high degree of immunity, on others a weak one. Second, it has to derive all justified beliefs from this terminus. Theories differ, likewise, on how beliefs are to be derived from those foundations—inferentially, evidentially, in combination, or some other way. Finally, it has to apply the preceding to the agent; it has to explain how the knowing subject exemplifies justification. In other words, it has to explain whether subjects must have or show justification, and whether justification requires internal access or not.

Differences aside, however, all foundationalist epistemic theories are committed to a conception of justified belief as having a hierarchical structure (of some sort) in which basic cognitions (however conceived) stop a regress of justification (however formulated), while non-basic cognitions derive justification mediately by way of regress-stoppers. Every justified belief finds some place within this structure; nothing falling outside of it is justified.

These common commitments serve to differentiate foundationalist theories from their rivals. Considerations of coherence, social context, reliability, and epistemic virtue may well play important roles within a foundationalist theory, but they do so only subordinate to foundationalist strictures. A foundationalist structure of justified belief
will typically be coherent; coherence can serve as a source of justification so long as it doesn’t violate the foundationalist ban on circular justification.² Again, a foundationalist account may involve reference to contextualist forms of justification, but only insofar as doing so bears in some way on solving the regress problem on foundationalist terms.

In any case, within these constraints, identifying an epistemic theory as “foundationalist” is, if not quite a trivial exercise, still a relatively straightforward affair. Whatever their other differences, most epistemologists would agree that the foundationalist project begins with Descartes’ Meditations (granting its distant roots in Aristotle’s Posterior Analytics) and continues in modern and contemporary epistemology in the work of Russell, C.I. Lewis, Chisholm, Alston, Moser, Audi, Fumerton, and others.³

Given the preceding account, the following definitory or taxonomic claims turn out to be nearly trivial truths:

1. Every foundationalist theory must answer the regress problem.

2. The structure of an epistemic foundationalist theory turns primarily on how one characterizes the regress problem, and how one construes the properties of the justificatory relation.

3. Foundationalism, coherentism, and contextualism are distinct solutions to the regress problem. No foundationalist theory is an instance of coherentism or contextualism; no coherentist or contextualist theory is an instance of foundationalism.⁴

It’s no exaggeration to say that these trivial truths define the basic problematic of contemporary epistemology. No serious conception of epistemic justification gets off the ground without addressing itself to the regress problem, and either accepting the
foundationalist solution to the problem, producing a different solution, or rejecting the
problem and explaining why. So understood, foundationalism is a clearly articulated
problem with a clear set of solutions.

0.3 Conceptualizing Foundationalism about Ethics

It’s a little-discussed truism that the terms “foundation” and, occasionally,
“foundationalism” also find expression outside of epistemology. One finds talk of
“foundations” in the philosophy of mathematics, philosophy of science, in ethics and
political philosophy, and in disciplines outside of but in some ways connected to
philosophy, e.g., anthropology, sociology, political science, literary theory, jurisprudence,
cultural studies, and so on. Sometimes reference to “foundations” in these contexts is
relatively cursory, casual, or rhetorical, intended only as a rough synonym for “basic” or
“indispensable” knowledge. But sometimes writers use “foundation” and its cognates
more self-consciously, to convey a more systematic attachment to a distinctive theoretical
conception. It is, however, often unclear how foundationalism in this latter sense is to be
understood, and particularly unclear how it relates, if at all, to epistemic foundationalism.

These issues arise in a particularly acute way when we come to “the foundations
of ethics.” Here as elsewhere, there are times when “foundations”-talk is used too
casually to be taken very seriously. But sometimes it clearly is meant to be taken
seriously. A number of philosophers have offered fairly elaborate analyses of, or
polemics against, what they call “ethical foundationalism.” Some of these analyses
explicitly use the term “ethical foundationalism” (or “foundationalism” in an obviously
ethical context), while others use the phrase “foundations of ethics” in a way that at least
implicitly invites the inference that anyone who defends that idea is in some sense an ethical foundationalist. And yet it is for the most part astonishingly unclear what any of these writers are talking about. Claims about the “foundations of ethics” are all over the map, and all over it in ways that evince no awareness of the difficulties that arise from the unclarity of their being so. What presents itself as a clearly defined issue in epistemology becomes, when transferred to ethics, a morass of confusion.

Since there is no canonical doctrine called “ethical foundationalism,” and no standard literature glossing the concept, anyone intending to use it for some non-casual theoretical purpose is obliged to fix its meaning. There are, broadly speaking, two distinct ways of going about this—one lexicographically driven, the other theoretically driven. On the lexicographic strategy, we gather instances of foundations-talk in ethics and figure out what the instances have in common. The result, ideally expressed in a biconditional, is our definition of ethical foundationalism. On the theory-driven strategy, by contrast, we forget about how the term is actually used in ethics, and look to model ethical foundationalism on its epistemic cousin. Since epistemic foundationalism is a response to the epistemic version of the regress problem, ethical foundationalism must be the ethical version of the same problem.

As it happens, neither strategy quite succeeds. The lexicographic approach is a complete failure: there is in fact no way “to gather instances of foundations-talk from the literature and figure out what they have in common.” But the theoretical approach is at best a mitigated success: there are, as we’ll see, more pitfalls in “modeling” ethics on epistemology than might be surmised.
0.3.1 The Lexicographic Strategy

Suppose we take a somewhat literal-mindedly terminological view of the matter and adopt the view that the key to understanding foundationalism in ethics is to understand *foundations-talk in ethics*. On this view, a philosopher is a foundationalist about ethics if and only if he uses the term “foundationalist” to describe his commitments about ethics. Thus, anyone who calls himself a foundationalist is one. The task of understanding ethical foundationalism on this view would be to grasp the doctrine common to all and only those who engage in “foundations”-talk in ethics.

Obviously, this approach won’t work precisely as stated. Since, as I’ve already noted, some philosophers use “foundations”-talk too casually to denote a commitment to anything in particular, “foundations”-talk will not, as such, be a *sufficient* condition for a commitment to foundationalism about ethics.

It won’t be necessary, either. After all, some philosophers, while not explicitly invoking the term “foundations,” discuss notions that are at least close cognates of that idea, e.g., “Archimidean Points,” “first principles,” “ultimate principles,” “axioms,” “sources (or bases) of morality (or norms, or normativity),” and so on. If these cognates express ideas that are close *enough* to foundationalism, then a philosopher’s use of the term “foundationalism” will not be a necessary condition of his being a foundationalist. He could use a different term and still be one.

Clearly, our original formulation needs some patching up. Suppose then that we adopt the view that non-casual foundationalist-talk and non-casual cognate-foundationalist-talk is a *sufficient* condition for a commitment to ethical foundationalism.
This might seem an initially promising approach: it neatly evades both of the preceding two problems. But of course, in doing so, it merely relocates them. Casual foundations-talk is a problem for our necessary condition, but it can’t be disposed of by adding a “non-casual” rider: we can’t distinguish casual from non-casual instances of foundations-talk unless we first have in hand a criterion of what makes foundations-talk foundationalist. Cognate foundations-talk is a problem for our sufficient condition, but it can’t be disposed of by adding a “cognate-talk” rider: for how precisely do we identify what counts as a non-casual cognate of foundationalism unless we can compare it to some independent criterion for what counts as an instance of foundationalism? In any case, even if we (somehow) succeeded at both tasks, there’s no reason assume that the assembled set of non-casual instances of foundationalism and its cognates would share anything in common. Obviously, if we need a criterion in hand in order to identify the relevant cases of foundations-talk, the talk cannot provide the criterion; it presupposes one.

Even if somehow the lexicographic approach could surmount the preceding difficulties—and it couldn’t—it would likely purchase them at the price of philosophical parochiality. For it should be obvious that a lexicographical approach will confront colossal complexity the minute we move beyond contemporary English-speaking philosophical discourse to discourse in other historical contexts or languages. If we do, we introduce questions about translation, context, and commensurability that vitiate the task of finding either necessary or sufficient conditions by a lexicographic approach.

To apply a lexicographic approach to non-English texts, we need to know which terms in the text are best translated by the English term “foundationalism.” But that isn’t
in the least obvious. In the Republic, Plato makes reference to tou agathou idean (“the Form of the Good”) and pantos archein ion (“the unhypothetical first principle of everything”). In the Nicomachean Ethics, Aristotle makes reference to t’agathon kai to ariston (“the good and the best end”), t’anthropinon agathon (“the good of man”), eudaimonia (“happiness,” or “flourishing”), and proton arche (“first principle”). In De Officiis, Cicero offers a brief justification for translating Greek philosophy into Latin, and translates Aristotle’s arche into Latin as extremum et ultimum bonurum (“the final and ultimate good”), along with summum bonum (“chief good”—a practice more or less followed in Aquinas’s discussions of the same topics in the Summa Theologiae and in his Commentary on Aristotle’s Ethics. Meanwhile, the same concepts are Arabicized in Al Farabi’s Tahsil al-Sa’adah (Attainment of Happiness), the first part of his Islamicized conspectus of the philosophies of Plato and Aristotle, itself the basis of much Islamic philosophy. Kant’s major work on ethics is a Kritik (Critique), his major discussion of meta-ethics is a Grundlegung (Groundwork or Foundation), and the fundamental concepts in his ethical system are Grundsatz (“principle”) and Imperativ (“imperative”).

Since it’s not clear which of these terms is best translated as “foundation,” it’s unclear which signifies a commitment to foundationalism. A fortiori, it’s unclear whether the use of any of them is either necessary or sufficient for a commitment to foundationalism. In fact, it’s unclear whether they bear any relation at all to foundationalism, or, for that matter, any relation at all to one another.

It seems obvious that we gain little insight into the character of “the foundations of ethics” by focusing narrowly on foundations-talk or even cognate foundations-talk, or non-casual cognate foundations-talk, etc. no matter how many qualifications we stick into
our definition. The relevant issue is the criterion for a theory’s being foundationalist, not foundations-talk as such—which pushes us to the theory-driven approach.

0.3.2 The Theory-Driven Strategy

Suppose that we dispense with the lexicographic approach and consider the matter from a more broadly conceptual or theory-driven perspective. We might, by analogy with the epistemic case, call an ethical theory foundationalist just in case it’s a justificatory theory about ethics with a double-tiered structure corresponding by analogy to what epistemologists would call a foundation and superstructure. And the most obvious way of cashing out the analogy is to do so in terms of the infinite regress problem at the center of epistemic foundationalism. As John Post puts the point, “[r]egress arguments are not limited to epistemology”; they “serve … many masters, from epistemology to ethics to metaphysics….” Though all regress arguments share a common form, “the same argument form can be instantiated by different subject matters, of which epistemology is but one.”

The issue then is not foundations-talk per se, but commitment to a particular conception of justification in ethics, regardless of the terminology used to describe it.

Plausible as this proposal may seem, however, it gives rise to three problems present but little discussed in the literature. (1) For one thing, since the subject-matters of epistemology and ethics overlap in complex ways, when it comes to ethics, the regress argument isn’t merely “instantiated by different subject matters,” but multiply instantiated within the same subject matter. In one sense, “the” regress problem motivates an epistemic approach to the foundations of ethics; in another sense, the “same” problem motivates a meta-ethical approach to the same subject. It is thus unclear
how many distinct regress problems are in play here, and how they relate to each other.

(2) If we adopt the epistemic approach to the foundations of ethics, it is safe to say that while the question is clear, the candidate answers are not. (3) If we adopt the meta-ethical approach to the foundations of ethics, the reverse of (2) seems to be the case: the answers are clear, but the question is not. I take these three issues in turn below.

0.3.3 Multiply-Instantiated Regress Arguments in Ethics

Foundationalism is motivated by “the” regress problem. In the epistemic case, the regress concerns the epistemic justification of belief. Suppose now that we want to model an ethical form of foundationalism on this notion. There are two distinct but easily conflated ways of doing it.

One way is to treat ethical foundationalism as a special case of epistemic foundationalism. In this case, we adopt epistemic foundationalist strictures about justified belief generally, and apply those strictures to the case of justified moral belief. The foundation of ethics on this view is an epistemic foundation: what justifies moral claims is what epistemically justifies them. It is thus misleading in this context to say, as Post does in the quotation cited above, that “the same argument form [i.e., the infinite regress form] can be instantiated by different subject matters of which epistemology is but one” and ethics another. If we take the epistemic approach to the foundations of ethics just described, there is one regress problem leading to one foundationalist solution: there is one foundation for all of justified belief, whether moral or non-moral. It’s not that there are two distinct regress problems, one epistemic and one ethical, with two distinct solutions, epistemic and ethical. It’s that the epistemic regress problem has a single
epistemic solution, but that the solution has a variety of applications, each differentiated by the content of the propositions being justified. Scientific beliefs are justified differently from ordinary perceptual beliefs, which are themselves justified differently from, say, moral beliefs. So epistemic foundationalism leads us to an epistemic approach to the foundations of ethics.

A second way of modeling ethical foundationalism on epistemic foundationalism is to treat the two as (otherwise unrelated) analogues of each other. In this case, we note the formal motivation for and structure of epistemic foundationalism; we then apply this entirely formal structure to ethics while abstracting from its specifically epistemic content. As we’ve seen, epistemic foundationalism gets its motivation from the regress problem: foundationalism is one of several solutions that arises from treating the idea of an infinite regress of justifications as a reductio ad absurdum. And, as a matter of structure, epistemic foundationalism tells us that justified belief has a two-tiered structure: the foundation constitutes the terminus of justification, while supporting a superstructure of justified beliefs “atop” it.

To get an ethical analogue of this view that is not merely a special case of epistemic foundationalism, we formulate a specifically ethical version of the infinite regress problem, not reducible to the epistemic version. To do this, we need a conception of ethical justification that is itself not reducible to the epistemic justification of ethical beliefs. We then run this regress argument, as we would in epistemic contexts, as a reductio—bearing in mind, however, that the reductio in question is specifically ethical, not epistemic. And we end with an ethical analogue of the two-tiered conception of justification we find in epistemology: a foundation of ethically justified claims
asymmetrically supporting a superstructure of them. In this case, epistemic foundationalism leads us to a non-epistemic analogue of an epistemic approach to the foundations of ethics. Call this the meta-ethical approach to the foundations of ethics. Notice that by contrast with the epistemic approach, the meta-ethical approach to the foundations of ethics involves two distinct (but analogous) regress problems.

The distinction between epistemic and meta-ethical approaches to the foundations of ethics is rarely if ever made clear in the literature. What we find instead is a confusing series of equivocations about whether a given discussion of the foundations of ethics is exclusively about meta-ethics or exclusively about moral epistemology or about both at once. A classic example is Rawls’s discussing intuitionism as though it contrasted principally with classical utilitarianism. Though Rawls takes the contrast between the two theories for granted—as do many textbooks—it is hardly clear that intuitionism is a theory of the same type as utilitarianism. In my terms, intuitionism is a foundationalist moral epistemology; utilitarianism is a foundationalist meta-ethics. Given this distinction, Rawls’s contrast is misplaced: we can differentiate theories that differ in some respect given a common subject matter, but not theories about entirely different subject matters. Arguably, the intuitionist/utilitarian contrast falls into the latter category.¹²

The same issues arise in (or sometimes derive from) work in the history of ideas. Historians of ethics, and ethicists in historical moods, have described Plato, Aristotle, Abelard, Aquinas, Hobbes, Locke, Spinoza, Kant, Mill, Herbert Spencer, and Moore as foundationalists about ethics. But the literature exhibits remarkable inconsistency and unclarity about what it means to be a foundationalist so construed. For every assertion that “X is a foundationalist,” it seems, one finds a dissenting voice in the literature.
asserting just the reverse about the same figure. In a large proportion of these cases, I would argue, the dissent over “X’s being a foundationalist” arises from scholars’ failure to distinguish X’s commitments to epistemic versus meta-ethical approaches to foundationalism.\footnote{13}

It would be easy enough to dismiss such disagreement as par for the course in scholarship: scholars can manage to disagree about almost anything, and a good deal of historical discussion is explicitly about “shadow history,” not history as such. But there is also the possibility that the scholarly disagreement arises from and feeds into the taxonomic confusion I’ve just described, in which case we have a sort of reciprocal process by which confusion is leveraged and multiplied: confusions in historiography abet confusions in contemporary discussion, which abet further confusions in historiography, and so on. Surely historiography has come to a point of diminishing returns when historians have trouble deciding whether Aristotle or Kant are paradigm instances of foundationalism or anti-foundationalism—at the very time when both Aristotle and Kant have assumed center-stage in contemporary ethical discussion.\footnote{14}

It may seem sufficient to distinguish the epistemic from the meta-ethical approaches to the foundations of ethics and leave the matter there, on the assumption that no equivocation should arise so long as we distinguish the two approaches and keep them distinct. But while the distinction prevents equivocation, the distinction itself gives rise to a series of puzzles requiring resolution.

We see the significance of these questions if we reformulate them slightly. If both the epistemic and the meta-ethical approaches to the foundations of ethics can be defended, then any justified ethical theory will, by its own foundationalist strictures,
appear to have *two* foundations—one epistemic and one meta-ethical. This is, I think, a philosophically significant result, as neglected in the literature as it is odd. For one thing, precisely because it is rarely if ever noted in the literature, it’s highly prone to equivocation. If foundationalism entails that any justified ethical theory has two foundations, we’re going to have to be careful to distinguish between them. Since philosophers seldom do, it is almost always unclear in the literature whether discussions of the foundations of ethics are exclusively about meta-ethics or epistemology or simultaneously about both. More subtly, it is unclear what implications claims about the epistemic approach have for the meta-ethical one (and vice versa): In what ways, if any, are commitments to epistemic foundationalism linked to meta-ethical foundationalism? Some philosophers insist that the two doctrines are related, differing on the relation. Others insist that there is no relation. Some take both doctrines to be true, others take both to be false, and others affirm one while denying (or ignoring) the other.

One common way of bypassing these questions altogether is to settle them by stipulation. Authors will characteristically tell us that “for present purposes,” or “in the present context,” they will treat, say, the epistemic approach *as though* it were unrelated to the meta-ethical one. But treating *x* as though it was unrelated to *y* doesn’t decouple the actual relations between them. If the falsehood of *x* entails the falsehood of *y*, it won’t help the case for *y* by *stipulating* that the truth of *x* “in the sense intended in the present discussion” has no bearing on a discussion of the truth of *y*.

This issue has special poignancy in a discussion of foundationalism. The whole idea of foundationalism is that justificatory claims in a certain domain terminate in a single unified foundation that confers justification on all justificanda of a certain kind
without needing justification of any kind. Suppose that now we confront the situation in which epistemic foundationalists argue for an epistemic foundation for moral claims, while meta-ethical foundationalists argue for a meta-ethical foundation for moral claims. Suppose further that the epistemologists simply refuse to discuss meta-ethics (it doesn’t “interest” them), while the meta-ethicists do the reverse. Recall that each set of theorists claims to be foundationalists about ethics. Consider the fact that the subject-matter of both theories is “claims about morality.” Now consider the studied insistence that the subject-matter of an epistemic approach to foundationalism not be distinguished too sharply from a meta-ethical one. Barring some argument to the contrary, this insistence is consistent with the possibility that there is some overlap between the subject-matter of a foundationalist moral epistemology and that of a foundationalist meta-ethics. The trouble is, this latter result is prima facie inconsistent with the basic insight of foundationalism. Foundationalism requires every given justificatory domain to have a single unified foundation. But the solution-by-stipulation view implies that ethics has two foundations. This wouldn’t be a threatening result if (but only if) we had in hand an argument to the effect that epistemic and meta-ethical foundationalism address different (and exclusive) domains. But that is precisely the sort of thing that requires substantive analysis of some kind. It can’t be resolved by stipulation.

0.3.4 The Foundations of Ethics as an Epistemic Project

Suppose that we conceive of ethical foundationalism in epistemic terms. In that case, we adopt epistemic foundationalism as a general view about the structure of justified belief, and assuming that some of our moral beliefs are justified, infer that those
beliefs, qua justified, are either foundational or superstructural. If all moral beliefs are
foundational, we explain what it is about them that makes them so. If all of them are
superstructural, we provide an account of the relation between them and the foundation.
And if some happen to fall into each category, we have to divide moral beliefs into the
foundational and the superstructural, and engage in both tasks. Stated in broad outline, it
seems clear and obvious enough. But if we turn to the literature, what seems “clear and
obvious” becomes just the reverse.

As a preliminary, we need to get clear on what counts as “the literature.” For in
fact, in discussing epistemic foundationalism, we confront (at least) two very different
literatures. There is, on the one hand, the literature of general (as opposed to moral)
epistemology, where epistemic foundationalism is discussed as a very general doctrine
about justified belief as such. There is, on the other hand, the literature of moral (as
opposed to general) epistemology, which focuses specifically on justified moral belief,
typically in abstraction from non-moral belief. Though the distinction between these two
literatures is rough, it is nonetheless real, and the differences between the two literatures
correspond to a rather odd set of lacunae in discussions of the subject.

Epistemic foundationalism, as I’ve discussed, is a theory of justification for belief
with applications to moral belief. Given this, we would expect general epistemologists at
least to mention this fact. But as a purely bibliographical matter, the fact is that they
rarely do. The standard examples in the literature are perceptual or scientific, not moral.
As far as the general literature is concerned, the implications of epistemic
foundationalism for justified moral belief are either too obvious to be discussed or else
not worth discussing or best left to those who specialize in moral epistemology.
If we move to moral epistemology, however, we don’t (as we might expect) find moral epistemologists picking up where the general epistemologists have left off, i.e., applying the general strictures of epistemic foundationalism to the specific case of moral belief. What we find instead is a quick inference from epistemic foundationalism to intuitionism: intuitionism is almost universally thought to have a monopoly on epistemic foundationalism as a moral epistemology. Oddly, however, the commitment to intuitionism is almost never depicted as following from—or even being connected to—any general commitment to epistemic foundationalism. Nor are there discussions of what would follow in moral epistemology (whether intuitionism or anything else) from a general commitment to epistemic foundationalism. It is in effect axiomatic in moral epistemology that intuitionism is the only foundationalist game in town, and that it is the only game in town without following from any more basic commitment to epistemic foundationalism about justified belief as such. More oddly still, intuitionism, though depicted as a form of foundationalism, is typically thought to be compatible with coherentism.

The disconnection between the two literatures leaves several questions essentially unanswered in the literature. For one thing, it is unclear what it would be to apply general foundationalist strictures to the case of justified moral belief. Second, it is unclear whether the application of these strictures would yield intuitionism or something else. Third, it is unclear whether or not intuitionism as defended in moral epistemology implicitly presupposes a commitment to foundationalism as a general epistemology. Fourth, it is unclear how the incompatibility of foundationalism and coherentism in general epistemology can be reconciled with the putative combination of intuitionism and
coherentism in moral epistemology. Without answers to these questions, we lack a viable way to give sense to the idea of the foundations of ethics as an epistemic project.

0.3.5 The Foundations of Ethics as a Meta-Ethical Project

Suppose instead that we conceive the foundations of ethics in meta-ethical terms. In this case, the literature suggests that the project of discovering the foundations of ethics is to be conceptualized in terms of the distinction between consequentialist and deontological theories.\(^\text{15}\)

Between say the 1930s and the 1970s, the taxonomic principle governing ethics was essentially meta-ethical. Theories were classified according to their meta-ethical structure, i.e., by what they took to be “meta-ethically basic.” And so ethicists devised (what they took to be) two jointly exhaustive and mutually exclusive classificatory options, thought to account for everything worth accounting for, from Plato’s *Republic* to Frankena’s *Ethics*: namely, “consequentialism” and “deontology.”\(^\text{16}\) Consequentialists took states of affairs as basic, while deontologists took acts as basic. Consequentialism thus instructs us to maximize intrinsically valuable states of affairs; deontologists enumerate lists of obligatory, impermissible, and supererogatory actions.

Since those two tasks were thought to exhaust all the relevant conceptual space, the task of the foundations of ethics in this taxonomic context was either consequentialist or deontological. Though philosophers only rarely used this terminology, the consequentialist/deontological distinction did in principle undergird and motivate a foundationalist enterprise for ethics. *Consequentialist foundationalisms* singled out a basic duty to maximize intrinsic value, treating this as the foundation of ethics, and then
generated a superstructure of prescriptions that did so. *Deontological foundationalisms* identified a basic non-consequentialist duty, treating this as the foundation of ethics, and then derived all subsidiary duties from this principle. Thus an argument for ethical foundationalism was either an argument for or against some variety of consequentialism or deontology.

While a remarkable amount of twentieth-century ethics takes the consequentialist-deontological *principium divisionis* for granted, it was clear as early as the 1950s and patently obvious by the mid 1970s that the division was seriously inadequate, and this for at least three major reasons.

There was, for one thing, no good definition of “consequentialism.” The term was coined by Elizabeth Anscombe in her 1958 essay “Modern Moral Philosophy” to mark a species of twentieth-century utilitarianism committed to denying the doctrine of double effect (DDE). On this view, one species of utilitarian (e.g., Mill) accepts DDE; the other (e.g., Sidgwick) rejects it. Some philosophers, however, use “consequentialism” not as a species of utilitarianism but as a synonym for it, so that “consequentialism” and “utilitarianism” are essentially interchangeable, and both deny DDE. A third group of philosophers takes “consequentialism” to denote a higher-order feature of utilitarianism, as in J. L. Mackie’s claim that we “can retain the consequentialist structure of utilitarian theory,” but replace utilitarian goals with different ones. 17 This view has more equivocal implications for DDE.

The dueling terminology of “consequentialism” and “teleology” adds a new series of complications to the preceding set. Thus Frankena describes “teleology” as the genus of which utilitarianism and ethical egoism are species. Meanwhile, Rawls (claiming to
follow Frankena) excludes egoism but adds Aristotelian and Nietzschean perfectionisms to the list of teleological theories. All in all, it is simply unclear what is being referred to when philosophers make reference to “consequentialism.” If we take “deontology” to be the denial of consequentialism, it is a fortiori unclear what it means as well.

Second, there was no generally accepted account of whether consequentialism and deontology were meant to be exclusive categories of theory, or whether consequentialist-deontological hybrids were possible (in which deontic elements functioned as “side-constraints” on consequentialist maximizing). The possibility of consequentialist-deontological hybrids undermines the idea that either consequentialism or deontology could function as a foundation for ethics. A hybrid view is typically pluralist: both consequentialist and deontological considerations account, separately, for moral norms. The adoption of this sort of pluralism, however, is ipso facto a rejection of the idea that there is one source of normativity—hence a rejection of the idea that there is a single foundation underlying a unified structure of prescriptions, hence of the meta-ethical version of the foundationalist project for ethics.

Third, some philosophers have protested that the consequentialist/deontological schema is non-exhaustive. As Jan Narveson argued four decades ago, the schema fails to account for contractarian theories. As John Cooper argued three decades ago, the schema fails to account for virtue-based theories of an Aristotelian variety. After Rawls’s turn to constructivism two decades ago, it should have become equally clear that the schema fails to account for constructivist versions of contractarianism, or constructivism as such.
Not being exhaustive, the consequentialist/deontological schema had no way of accounting for ethical projects that were neither consequentialist nor deontological but apparently foundationalist (e.g., Aristotle’s), and little to say about projects that were neither consequentialist nor deontological nor apparently foundationalist (e.g., Rawls’s or Gauthier’s). In the absence of an argument demonstrating that such theories had to be either consequentialist or deontological, the distinction could not serve as a particularly robust motivation for a foundationalist project in ethics. Thus for meta-ethical foundationalism to be a viable project, we need either a way of patching up the consequentialist/deontological distinction or a new rationale for foundationalism.

0.4 The Plan of the Present Study

As remarked above, my aim in the present study is three-fold: to defend a particular conception of the problem that goes by the name “the foundation of ethics,” to offer a sketch of the solution, and to offer an extended critique of the neo-Kantian solution offered by Christine Korsgaard in *The Sources of Normativity*.

Conceiving the problem, as I see it, is principally a matter of getting straight on the similarities, differences, and relations between epistemic and meta-ethical varieties of foundationalism. I begin in Chapter 1 by asking in meta-philosophical terms about the possible relations that epistemic and meta-ethical foundationalisms *could* bear to one another, which I call *homonymy, analogy, reductionism, and a reductionist-analogy hybrid*. Homonymy is the view that the two projects are utterly unrelated to each other. Analogy (in the non-hybrid sense) is the view that they are related *only* by analogy, and by no more than that. Reductionism is the view that one is a deductive consequence of
the other. And the reduction-analogy hybrid is, as the name suggests, an attempt to marry
a weaker version of the analogy view to a modified form of reductionism. This meta-
philosophical inquiry suggests the possibility of what I call a *comprehensive normative
foundationalism*—an approach to the foundations of ethics that not only includes both
epistemic and meta-ethical components, but does so in a way that is both coherent and
consistent with foundationalist strictures.

In Chapter 2, I discuss the *epistemic* component of this comprehensive view,
identifying its rationale and sketching the argument for it. Throughout the twentieth
century, philosophers have supposed that a commitment to an epistemic approach to the
foundations of ethics should entail intuitionism. I dispute this claim. Contrary to much
discussion in moral epistemology, I argue that an epistemic approach to the foundations
of ethics is best seen as a particular application of a general commitment to epistemic
foundationalism. For a variety of highly contingent reasons, philosophers have failed to
note the complexity of this application. In particular, they have paid insufficient attention
to non-doxastic versions of epistemic foundationalism (i.e., versions of foundationalism
committed to non-doxastic foundations), and have paid almost no attention whatsoever to
the implications of non-doxastic foundationalisms for *moral* epistemology. In laying out
the logic of the application of epistemic foundationalism to moral contexts, I open logical
space (among other things) for what I call a *non-intuitionist empiricist foundationalism*
of basically Aristotelian provenance.

In Chapter 3, I turn to a brief discussion of the meta-ethical component of
comprehensive foundationalism, sketching its structure and identifying its varieties. The
aim here is to offer a rationale for meta-ethical foundationalism that supersedes the sort
based on the distinction between consequentialism and deontology. The rationale, I suggest, lies with the so-called “Why be moral?” question. Meta-ethical foundationalisms are all attempts to answer this question, differentiated by how they interpret it, how they answer it, and what they build atop the answers they give.

In Chapter 4, I sketch the features of the most fundamental challenge to meta-ethical foundationalism, namely, the claim that unlike epistemic foundationalism, it constitutes an attempt to answer a pseudo-question, viz., “Why be moral?” Philosophers have often thought this question either pointless or illegitimate. In the chapter I offer the (putatively) best argument for thinking so—what I call the “erotetic dilemma.” In Chapter 5, however, I respond to the erotetic dilemma by identifying its flaws, vindicating meta-ethical foundationalism, and reformulating the version of the “Why be moral?” question in order to avoid the dilemma.

In Chapter 6, I pull together the claims of the preceding chapters to offer a positive account of the comprehensive foundationalism I myself espouse—an empiricist foundationalism in moral epistemology, and ethical egoism in meta-ethics. I end the chapter with a meta-philosophical account (based on the framework introduced in Chapter 1) of how the two commitments relate to one another, and generally how epistemic and meta-ethical components of a comprehensive foundationalism fit together in a way that makes sense and is consistent with foundationalism as such.

As I argue in Chapter 2, egoism as a meta-ethical foundationalism contrasts with—and constitutes a rival to—divine command theories, utilitarianism, contractarianism, and Kantian deontology. After an “Interlude” between Chapters 6 and 7, I offer an extended critique of a rival foundationalism of a Kantian variety. Using the
framework developed in the preceding chapters, I suggest that Korsgaard’s Kantian foundationalism fails—and that the explanation for its failure lies precisely with what my own comprehensive foundationalism—empiricism and egoism—answers. Chapter 7 identifies the difficulties in Korsgaard’s formulation of “the normative question”; Chapter 8 takes issue with her answer to it.

The study ends with two appendices keyed to aspects of the discussion of Korsgaard, one discussing Rawls’s (and by implication Korsgaard’s) views on the relation between constructivism and realism, the other discussing reconstructions of Korsgaard’s argument for the source of normativity.  

0.5 Notes


For non-doxastic accounts of foundationalism, see David Kelley, The Evidence of the Senses: A Realist Theory of Perception (Baton Rouge: Louisiana State University Press, 1986); David Kelley, “Evidence and Justification,” Reason Papers, vol. 16 (Fall 25

2 For useful discussions, see Kelley, *Evidence of the Senses*, ch. 6, and Susan Haack, *Evidence and Inquiry*, ch. 6. I thank Michael Young for helpful discussion on this issue.


Haack claims to offer what she calls a “foundherentist” view that blends elements of foundationalism and coherentism, but I think Wallace Matson successfully argues that Haack’s view is ultimately foundationalist. See Wallace Matson, Review of Susan Haack’s *Evidence and Inquiry*, *Reason Papers*, vol. 22 (1997), pp. 161-63.

5 I divide these roughly by decade and list them in roughly chronological order; the list is far from exhaustive.


6 Though the two approaches might well converge on the same result, they are, as approaches to the subject, mutually exclusive.


8 Plato, Republic, 508e, 511b.

9 Aristotle, Nicomachean Ethics, 1094a22, 1094b7, 1097b20, 1098b2.

10 Cicero, De Officiis, I.2, I.9; cf. Aquinas, Summa Theologiae, I-II, Q. 1-5, as well as the relevant sections of Aquinas, Commentary on Aristotle’s Nicomachean Ethics. For extensive discussion of the move from Aristotle to Aquinas via Cicero, see Alasdair MacIntyre, Whose Justice? Which Rationality? (Notre Dame: University of Notre Dame Press, 1988), chs. 6-11. My discussion in this section owes a great deal to this book.


12 Rawls, Theory of Justice, secs. 5-7.


A newer literature focuses on “reasons for action,” but for purposes of this study, I ignore the possibility that it might be useful for an inquiry into the foundations of ethics. For one thing, the literature on reasons for action typically presupposes a sort of value pluralism that is incompatible with foundationalism. So while it may be of use for discussing anti-foundationalism, it cannot serve as a motivation for foundationalism. Second, to put the matter bluntly, I find the subject matter of the literature question-begging and obscure. It is simply unclear (to me) what a “reason for action” is supposed to be, why the use of the term is necessary, what one is committed to in using it, and what problems are supposed to arise for a view that bypasses it. Contrary to the assumptions of the literature, as far as I can see, the one topic distinctive of the “reasons for action” literature—internalism versus externalism—can in fact be discussed without reference to the locution “reasons for action” (as I do in chapters 5-6). Likewise, the so-called “Humean theory of motivation,” often described as a “dogma of philosophical psychology” is as typically stated too unclear to be defended or rebutted.

In my view, virtually every contemporary interlocutor in the “reasons for action” debate is, in my terminology, either a foundationalist or an anti-foundationalist, and if a foundationalist, categorizable by one of the meta-ethical categories I discuss in Chapter 3 of the present study. I thus discuss such authors in my own terms rather than in those current in the literature.


The preceding two paragraphs come from my review of Stephen Darwall’s *Consequentialism, Teaching Philosophy*, vol. 28, no. 3 (September 2005), p. 283.


The classification of Rawls’s theory is a vexed issue. Arguably, Rawls’s constructivism is a form of Kantianism, hence of deontology, but it is equally a form of contractualism. But if contractualism is not a form of deontology (see note 20 above), then neither (it would seem) is Rawls’s view. Since Rawls’s discussion of the matter is obscure at best, it is possible that Rawls’s view may well turn out to be deontological or a hybrid view. In the first case, his theory would not be a counterexample to the consequentialist/deontological schema. In the second case, his theory would show its non-exclusivity, not its non-exhaustiveness. But the very unclarity of how to classify Rawls’s theory lends some plausibility to the non-exhaustiveness objection.

23 I’m grateful to Carrie-Ann Biondi and Michael Young for extensive comments on and criticism of this chapter.
CHAPTER 1:

FOUNDATIONALISM, EPISTEMIC AND META-ETHICAL:
A TAXONOMY OF CONCEPTUAL POSSIBILITIES

1.1 Introduction

“Epistemic foundationalism,” as we’ve seen, denotes a single familiar doctrine in contemporary analytic epistemology, with some (uncertain) implications for ethics. Meanwhile “meta-ethical foundationalism,” a less familiar phrase, refers in a generic way to any non-epistemic justificatory account of moral norms with a double-tiered structure whose foundation terminates some regress of justifications, and whose superstructure sits atop it.

It’s tempting to suppose that we can simply make the preceding distinction and leave the matter there. If we do, it seems, we have two different doctrines governing two different subjects, related to one another by little more than a loose but philosophically uninteresting analogy. Having done so, we’ve established or identified a division of theoretical labor between epistemology and meta-ethics, and we’re entitled to move on. The epistemologists are free to do epistemology, the meta-ethicists are free to borrow epistemic metaphors as they please, and never need the twain meet.

This temptation has, as I argued in the Introduction, structured assumptions in epistemology, moral epistemology, and meta-ethics. While epistemologists have recently
begun to “resurrect” foundationalism as a general theory of epistemic justification, they rarely if ever discuss its implications for justified moral belief. At the same time, moral epistemologists who discuss foundationalism in moral contexts focus almost exclusively on intuitionism, and do so while bracketing the relation between intuitionism as a moral epistemology and foundationalism as a general theory of epistemic justification.

Meanwhile, discussions in meta-ethics (and political philosophy) often borrow the metaphor of a foundation, but don’t clearly tell us how the metaphor is to be cashed out. It is thus radically unclear how the general epistemic doctrine, its application to moral belief, and its meta-ethical analogue relate to each other, if they relate at all. Anyone reading the literature (or the three literatures) could easily come to the conclusion that epistemic foundationalism has no implications for justified moral belief, that epistemic foundationalism in moral epistemology bears no relation to epistemic foundationalism anywhere else, and that there is no connection between foundationalism in epistemology (whether as a general doctrine or as a specifically moral one) and foundationalism as a meta-ethical thesis. Unfortunately, our reader would find few arguments in the literature for any of these claims, either—so that the reverse of each of them could, for all the literature tells us, equally be true.¹

The impulse to “leave it at the distinction and move on” is, I think, little more than institutionalized incuriosity.² For one thing, it ignores the real possibility of equivocation. In distinguishing between epistemic and meta-ethical foundationalisms, it may seem at first as though we’re distinguishing between two clearly distinct items, one epistemic and one meta-ethical. But as we’ll see, there are more conceptual possibilities here than at first meets the eye, and it’s important to keep them straight.
The *laissez-faire* impulse also ignores the need for integration. In stipulating that two doctrines are unrelated, one courts the danger that they are in fact related, and that the stipulation to keep them apart merely serves to conceal this fact. Suppose *ex hypothesi* that epistemic and meta-ethical foundationalisms are related in some conceptually important way. If so, the stipulation to keep them apart conceals from view how the truth-conditions of the one doctrine are related to those of the other. If they are related, but we ignore the relation, we remain in the dark about the logical consequences of a commitment to the one or the other. What ought to be immediate inferences—the truth (or falsity) of say, the epistemic doctrine entails (or doesn’t entail) the truth (or falsity) of the meta-ethical one—become inferences never made. The stipulation also obscures the possibility that the two doctrines cohere with and mutually support one another. In that case, what may have seemed like two different theories may well, when conjoined, be two constituents of a larger theory made more powerful by the conjunction.

As it happens, I think all of the preceding are the case. The failure to get clear on the relation between epistemic and meta-ethical foundationalisms has in fact led to equivocations that run throughout the literature. Given these equivocations, it is unclear when we can make inferences from epistemic to meta-ethical contexts, and when we can’t. For the same reason, the literature fails to recognize the possibility of a comprehensive foundationalist theory incorporating epistemic and meta-ethical components. *A fortiori*, it fails to recognize the possibility of giving a higher-order account of the relation(s) *between* these components, or of a maximally coherent theory involving all *three* components: epistemic, meta-ethical, and the higher-order account that unifies them.
In what follows, I'll suggest that epistemic and ethical foundationalism can relate to one another in four basic ways which I’ll call homonymy, analogy, reductionism, and the analogy-reduction hybrid. (I stress that what follows is an enumeration of conceptual possibilities, not a definitive statement of how the two theories must relate.) Homonymy is the view that epistemic and ethical foundationalism are names for two irreducibly distinct and unrelated doctrines bearing no logical relations to one another. Analogy is the view that epistemic and ethical foundationalism are analogous but irreducibly distinct doctrines related only by similarity. Reductionism is the view that one doctrine is a mere deductive consequence of the implications of the other, so that the truth (or falsity) of one trivially entails the truth (or falsity) of the other. The analogy-reduction hybrid is a perhaps counter-intuitive (but entirely coherent) view consisting of the conjunction of modified versions of analogy and reductionism. Surprisingly, few of the preceding views have been held in a conscious and explicit way that makes their implications entirely clear.

1.2 Homonymy

Homonyms are words that have the same sound but have different and unrelated meanings. Homonymy, then, is the view that ‘epistemic foundationalism’ is related to ‘ethical foundationalism’ as any two homonyms are (in a Pickwickian sense) “related” to each other. That is to say, they’re not related. Consider the word “bank.” In one sense, ‘bank’ names a financial institution; in another, it names the side of a river. Just as there is no important relation between financial institutions and rivers, on homonymy, there is literally no important relation between epistemic and ethical foundationalism. What we
have in each case is a similar name for two distinct and unrelated doctrines. The two
terms might look like species of a common genus (“foundationalism”), and given the
common terminology, we might be tempted to suspect that they share something in
common, however abstract or attenuated. But on homonymy, both appearances are an
illusion. There is no similarity to be discerned.

Homonymy is of course compatible with every possible combination of truth-
values for both doctrines. Suppose that homonymy is the case. If so, surely both
epistemic and ethical foundationalism can be true; homonymy has no bearing on that,
either way. But if they’re both true, the truth of one will be irrelevant to that of the other:
each will be a separate and distinct theory that happens to share the same name and
happen to be true. So it will be in the case where both doctrines are false, or the cases in
which one is true and the other false. This implies, in turn, that homonymy disables
inferences from the truth or falsity of one doctrine to the truth or falsity of the other. If so,
a refutation of epistemic foundationalism will not rebut ethical foundationalism, or vice
versa. Likewise, a defense of epistemic foundationalism will not tend to support ethical
foudationalism or vice versa. Each doctrine would have to be argued for or rebutted on its
own, and stand or fall on its own.

1.3 Analogy

An analogy is an elliptical (often pairwise or n-wise) assertion of similarity. For
present purposes, I take ‘X is analogous to Y’ to be elliptical for ‘X is significantly similar
to Y along some dimension, despite various differences along other dimensions.’ Note
that X and Y can be analogous on this understanding of “analogy” so long as they are
significantly similar along some dimension d1 while being different along dimensions d2, d3, … dn. This remains the case even when the dimensions of difference are as significant as the dimension of similarity. Thus a pumpkin can be similar to a basketball in color and shape while differing in price, purpose, texture, and internal structure. An analogy is better or stronger the more dimensions of similarity it involves, the more significant the similarities along a given dimension, and the fewer significant dimensions of difference arise between the analogues.³

In the present context, (pure) analogy is the view that epistemic and ethical foundationalism are distinct but analogous commitments motivated by distinct but analogous considerations. Thus, both epistemic and ethical foundationalism are similar qua theories of justification. But if we assume that epistemic and ethical justification are distinct sorts of justification with distinct purposes and criteria, then while both epistemic and ethical justification may share similar structural features, neither is identical or reducible to the other. Similarly, both epistemic and ethical foundationalism may be motivated by similar versions of the regress problem, and both may solve that problem in analogous ways. But if (pure) analogy is the case, then in neither case will the one regress be identical or reducible to the other. There will instead be distinctively epistemic and ethical conceptions of justification (and contexts of justification), in each case with distinctive regresses to distinctive termini. Finally, since both views share the same structure, both face analogous issues concerning the derivation of justified items from the foundation. On an epistemic view, our beliefs are epistemically justified by—i.e., derived from—an epistemic foundation. On an ethical view, moral norms are ethically justified by or derived from an ethical foundation. But while these are similar issues, on (pure)
analogy, no matter how similar they are, they are ultimately distinct: epistemic derivations do not map onto ethical ones, or vice versa. (Throughout section 1.3, references to “analogy” will denote references to “pure analogy.)

Analogy shares with homonymy a *compatibility* with all four combinations of truth-values for epistemic and ethical foundationalism. Like homonymy, analogy can treat the commitment to epistemic or ethical foundationalism as separate and independent matters: since the doctrines are similar, but not conceptually linked, the weaker the analogy asserted, the less the truth of one doctrine is related to that of the other. Assuming a very weak analogy between epistemic and ethical foundationalism, analogy can lead to the same results as homonymy as far as truth-values are concerned. Epistemic and ethical foundationalism can both be true, both be false, or one true and the other false, while bearing so negligible a similarity to one another as to amount for all practical purposes to homonymy.

But analogy differs from homonymy in asserting a real (if attenuated) connection between the two doctrines. For $x$ and $y$ to be analogous is for $x$ and $y$ to be distinct but similar in various significant and relevant respects. And similarity is a relation (however strong or weak) whereas homonymy is the absence of one. For this reason, the stronger the analogy asserted to obtain between epistemic and ethical foundationalism, the more likely that the truth of one will bear on the truth of the other.

The strength of an analogy turns in part on *why* the analogues are analogous. Typically weak analogues are weak because they share superficial or happenstance similarities: e.g., the pumpkin and basketball of my example above. Strong analogues are strong because they possess some underlying trait that gives rise to and explains the
similarities they share. They’re similar in virtue of something that makes them similar. Sometimes this underlying basis is causal. For instance, the principle of natural selection causally explains the similarity of biologically analogous (and heritable) traits. So natural selection explains why leopards are similar to jaguars (and leopard spots similar to jaguar spots). Sometimes, however, the underlying basis is logical. Logical species of a common (logical) genus may be analogous because they share a common generic feature.

Suppose that epistemic and ethical foundationalisms are species of a common genus, *foundationalism*. Both, we might say, are theories of justification, and thus share the common features that all theories of justification have qua justificatory. Both involve regress arguments, offer similar solutions to the regress, and have similar structures. So long as we think that there is a theoretically significant second-order account to be given about the genus “foundationalism”—so long as the similarities in question are relevant and important and not outweighed by equal or greater dissimilarities—the analogy between the two species of foundationalism will be strong, as will the connections between their respective truth-values. If both are true, this will probably be because the underlying generic argument for foundationalism is true, and happens to apply both to the epistemic and ethical domains. Likewise, if both are false, this will probably be because the generic foundationalist argument somehow fails, hence fails twice over. If the analogy between them is a strong one, it will be unlikely that one is true and the other false.

Unlikely, but not impossible. If epistemic and ethical foundationalism are species of a common genus, and one happens to be true while the other happens to be false, we’d be facing the unusual situation in which the generic logic of foundationalism applied to
one doctrine but not the other. In other words, the regress argument might (say) apply in the epistemic case but (for whatever reason) not in the ethical (or vice versa). In a case like this, however, we’d want to know why the discrepancy arose: what differentiating factor accounts for why foundationalist strictures applied in one case but not the other? One obvious explanation for such a discrepancy might be that while our non-moral beliefs meet foundationalist strictures, none of our moral beliefs do. Note that the controversiality of this explanation is what makes it informative: if epistemic and ethical foundationalism are strongly analogous, we ought to be very puzzled about the case in which one is true and the other false. The puzzlement-dispelling explanation in this case is bound to be as controversial as the puzzle was puzzling.

1.4 Reductionism

A reductionist view is one that holds that one or the other form of foundationalism is basic and the other simply a deductive consequence of the first, so that the concepts of the one are ultimately reducible to those of the other. In this instance, we have two sorts of reduction differentiated by what is being taken as basic, and what derivative. In epistemic reductionism, epistemic foundationalism is taken as basic, and ethical foundationalism is derived from it. In ethical reductionism, ethical foundationalism is taken as basic, and epistemic foundationalism is an instance of it. I take these in turn.

1.4.1 Epistemic Reductionism

Epistemic reductionism is the view that ethical foundationalism is simply the application of epistemic-foundationalist strictures, taken as basic, to the case of justified
moral belief, construed as just another domain to which epistemic-foundationalist strictures might apply. Epistemic foundationalism, recall, is a global thesis about justified belief as such: the idea is that every justified belief falls into a single foundational/superstructural order, with each justified item in the superstructure linked by derivation to some element of a single, unified foundation. If ex hypothesi we assume epistemic foundationalism to be true, and assume that some of our moral beliefs are justified, foundationalist strictures ought automatically to apply to justified moral belief, and so it ought to follow that our justified moral beliefs will find some place somewhere in the foundation or superstructure of our justified beliefs.

The differences between this view and homonymy are too obvious to belabor, but the similarities and differences between reductionism and analogy are worth bearing in mind. The weak version of analogy is essentially equivalent to homonymy: if $x$ and $y$ are related by a weak analogy, they may as well be treated as though they were homonyms. The strongest version of analogy tells us that epistemic and ethical foundationalism are structurally similar views, and thus likely to have the same truth-value. But analogy is not identity: even the strongest version of analogy leaves epistemic and ethical foundationalism similar but ultimately distinct views. And as we’ve seen, analogy makes a common truth-value likely, not necessary.

This has a subtle and interesting implication that differentiates analogy from reductionism. Obviously, if epistemic and ethical justification are distinct, to be epistemically justified is never by itself to be ethically justified. And on analogy, they are distinct. That implies, in turn, that $S$’s being epistemically justified in believing an ethical proposition $p$—even when $p$ is a prescription—is distinct from $S$’s being ethically justified.
justified in taking the action prescribed by $p$. If analogy is the case, believing that $p$ is not by itself a reason for action.\footnote{That is what reductionism denies. Analogy and reduction agree in regarding epistemic and ethical foundationalism as linked. They differ not only in the \textit{strength} but also in the \textit{type} of connection asserted. So where analogy asserts two distinct forms of justification, epistemic reductionism asserts one: ethical justification just \textit{is} epistemic justification applied to the context of ethical beliefs. Where analogy asserts two distinct regresses to two distinct termini, epistemic reductionism asserts one regress and one terminus. The connection between epistemic and ethical foundationalism, then, is less a matter of offering a second-order account of the relation between them than drawing out the implications of epistemic foundationalism as applied to justified moral belief.

This affects the possible combinations of truth-values in each instance. Granting that we have justified moral beliefs, the truth and falsity of ethical foundationalism turns entirely on that of epistemic foundationalism. Suppose that epistemic foundationalism is true, and that some of our moral beliefs are justified. In that case, ethical foundationalism must be true. Conversely, suppose that ethical foundationalism is false. In that case, epistemic foundationalism has to be false. The case in which epistemic foundationalism is true and ethical foundationalism is false is possible only if no moral beliefs are justified. But if reductionism is true, there simply cannot be a case in which epistemic foundationalism is false and ethical foundationalism true; the supposition makes no sense. If epistemic foundationalism is false, it has no true implications. \textit{A fortiori} it has no true ethical implications. Hence on this supposition, a refutation of epistemic foundationalism will knock ethical foundationalism out of the running.}
These observations have an important implication for standard treatments of intuitionism in the literature. On the standard view, intuitionism is regarded as a foundationalist moral epistemology, but treated as though it were possible, in principle, to detach a commitment to intuitionism in moral epistemology from foundationalism about justified belief as such. Some intuitionists, like Moore (and possibly Audi)\(^7\), may well be foundationalists about both justified belief and justified moral belief, and regard the former commitment as the basis for the latter. But these intuitionists are the exceptions that prove the rule. By and large, contemporary intuitionists (e.g., Prichard, Ross, Dancy, Williams, Nussbaum, arguably Rawls)\(^8\) have taken intuitionism to be a *sui generis* commitment in moral epistemology, and have either been neutral about foundationalism as a general epistemic position, or have rejected it outright in favor of coherentism or contextualism.\(^9\)

If epistemic reductionism is the case, these positions are untenable. If coherentism is true as a general epistemology, foundationalism is false. If foundationalism is false as a general epistemic thesis and reductionism true, then foundationalism cannot be true in moral epistemology. If intuitionism is a species of foundationalist moral epistemology (as most would say), then if reductionism is true, it cannot be detached from a general commitment to foundationalism or attached to coherentism. And if reflective equilibrium is coherentist (as most would also say—but not all\(^10\)), then if reductionism is true, it is incoherent to attach intuitionism to the method of reflective equilibrium. Given the popularity of this combination—i.e., (intuitionism and reflective equilibrium)—this is a significant finding. It implies that anyone committed to the preceding combination is
obliged explicitly to reject epistemic reductionism. As far as I can see, few philosophers have seen this, and few have argued for it.

1.4.2 Ethical Reductionism

Ethical reductionism is in effect the converse of epistemic reductionism. On this view, we take epistemic concepts (paradigmatically, “justification”) to be reducible to ethical concepts and take ethical justification to be basic. In this case, ethical foundationalism will be basic, and will entail epistemic foundationalism. This is not, I think, a view endorsed anywhere in precisely this form, but it is suggested by views expressed in the meta-epistemological literature by Firth, Chisholm, Alston, and others. Firth puts the point this way in a 1958 discussion of Chisholm’s *Perceiving: A Philosophical Study*:

The chief problems of perception are conceived by Chisholm in the traditional way, to be problems concerning the justification of perceptual judgments. We want to know, for example, under what conditions we can be said to have adequate evidence for our perceptual judgments, what form or forms the justifying arguments can take, and what reasons there are, if any, for thinking that these arguments are valid. For Chisholm, however, all such questions are questions of ethics, for to ask whether a judgment is justified, or whether we have adequate evidence for it, is simply to ask whether it is worthy of our belief. Epistemology is thus to be construed, in large part, as the ethics of belief.

This challenging thesis is a logical consequence of Chisholm’s analysis of the various terms he calls “epistemic”—terms such as “know,” “perceive,” “see,” “unreasonable,” “probable,” and “adequate evidence.” For when such terms are used epistemically they can all be defined, according to Chisholm, by using the ethical locution: “is more worthy of S’s belief than I,” where “S” may be replaced by the name of a person and “h” and “i” by names of propositions, statements, or hypotheses (pp. 4–5).

Thus ethical reductionism is the mirror-image of epistemic reductionism. Where epistemic reductionism takes epistemic concepts as basic and applies them to the case of
justified moral belief, ethical reductionism takes ethical concepts as basic and applies these to the epistemic analysis of justified belief as such. So S’s being justified in believing that \( p \) is an inherently ethical matter, so that non-evidential ethical considerations can trump epistemic ones when it comes to belief. Coming the other way around, epistemology plays no autonomous justificatory role with respect to ethics. There is not, for instance, an epistemic foundation that provides the basis for ethics. Ethics is instead the foundation for epistemology, and being the foundation, requires none of its own.

Ethical reductionism is thus an obvious basis for (and implicit presupposition of) intuitionism. Suppose that our intuitive ethical beliefs are justificatorily primitive. If so, they stop justificatory regresses without requiring further justification, and other normative concepts are built on or derived from them. In one sense, this is clearly a sort of foundationalism insofar as it posits a terminus of justification consisting of basic cognitions supporting a superstructure of claims justified by way of this foundation. But as a specifically ethical foundationalism, it bears a tricky relationship to epistemic foundationalism. In one sense, it is a species of epistemic foundationalism. It just happens to be one where ethical concepts are more basic than epistemic ones. In another sense, it is compatible with a rejection of epistemic foundationalism. If we adopt a version of epistemic foundationalism according to which ‘epistemic justification’ is distinct from ethical justification—if, for instance, epistemic justification grounds ethical claims—then ethical reductionism will reject epistemic foundationalism.

This, then, is what explains the apparent paradox of a reductionist non-foundationalist intuitionism. Foundationalisms, as we’ve seen, can either reduce the
ethical to the epistemic or the epistemic to the ethical. If the ethical is reduced to the epistemic, and we accept foundationalism, a commitment to intuitionism must be based on a commitment to a general form of epistemic foundationalism. If the epistemic is reduced to the ethical, and we accept foundationalism, a commitment to intuitionism can be epistemically primitive, and need involve no further commitment to a general form of epistemic foundationalism. Intuitionism is in fact consistent with a denial of a general epistemic foundationalism about justified belief. For if our ethical concepts are primary, they can in principle dictate our epistemic commitments. If ethics simply requires us to be coherentists, then our ethical commitments will override contrary epistemic commitments to, say, foundationalism.

As for truth-value combinations, then, we have a complicated situation. Suppose that ethical foundationalism is true and the relevant moral beliefs are justified. In that case, as we’ve seen, there is a sense in which epistemic foundationalism is true. Suppose now that ethical foundationalism is false. In that case, epistemic foundationalism could be true or false. The strange thing, then, is that ethical reductionism gives us truth-value combinations that end up being similar to homonymy.

The lesson here is that ethical reductionism is the view that coheres best with intuitionism. But the fact remains that few intuitionists have defended intuitionism that way.

1.5 Epistemic Reductionism and Analogy

I’ve so far been describing things as though epistemic reductionism contrasted with analogy. And in one sense, there is an obvious contrast to be drawn: the application
of foundationalist epistemic strictures to ethical beliefs (epistemic reductionism) is to be distinguished from a view that holds that there is a form of ethical justification that is analogous to but distinct from foundationalism (analogy). But could the two views somehow be made compatible with one another, and be combined in a single view? I think so. In fact, this is a genuinely interesting as opposed to merely nominal possibility. It also happens to be the view I shall ultimately defend in Chapter 6 (though it will take the intervening chapters to clarify what’s involved).

Analogy says that there are two analogous but distinct forms of justification, epistemic and ethical, where epistemic foundationalism is not reducible to ethical, or vice versa. We needn’t give epistemic foundationalism a special name in this case, but for ease of reference let’s call the relevant form of ethical justification *irreducible ethical foundationalism*. Suppose *ex hypothesi* that irreducible ethical foundationalism is true. In that case, there is a form of ethical foundationalism that is analogous to but irreducible to epistemic foundationalist strictures.

Now suppose that both epistemic foundationalism and epistemic reductionism are true, and that some of our moral beliefs are justified. In that case, there is a form of ethical foundationalism that is reducible to (derived from) epistemic foundationalist strictures. Call this latter thing *epistemically reducible ethical foundationalism*.

Obviously, irreducible ethical foundationalism is distinct from epistemically reducible ethical foundationalism: one is reducible to epistemic foundationalism, the other isn’t. But are they incompatible? Not necessarily. It could in principle be the case that both happen to be true. If epistemic foundationalism is true, and we have justified moral beliefs, then epistemic foundationalism will apply to ethical beliefs, and so
epistemically reducible ethical foundationalism will be true. But it might just be a separate fact about the moral world that irreducible ethical foundationalism—alogous but distinct from its epistemically reducible cousin—also happens to be true. In that case, we would have two distinct but analogous approaches to the foundations of ethics, one via epistemic foundationalism, one via some non-epistemic route. Indeed, we would have two foundations for ethics, one epistemic, one non-epistemic.

Recall that if analogy is the case, we may want a second-order account of why the analogy holds. If we conjoin analogy with reductionism, it makes sense to want a second-order account of why the conjunction turns out to be compatible. In other words, what we might want is not merely the conjunction of analogy and reductionism, but the conjunction of the two plus a meta-level account explaining the coherence of analogy with reductionism. The result would be a sort of comprehensive normative foundationalism, consisting of epistemic foundationalism plus its implications for justified moral belief, plus a separate analogous-but-irreducible form of ethical foundationalism, plus an account of how the total package coheres.

This rather abstract point is perhaps better conveyed by a particular example. Take, for instance, Aristotle’s view. Aristotle arguably espouses a general commitment to foundationalism at *Posterior Analytics* I.3 (hereafter *APo*). And so, we conclude, Aristotle is an epistemic foundationalist. This general epistemic commitment has implications for Aristotle’s moral epistemology as expressed, say, in the *Ethics, Politics,* and *Rhetoric;* his moral epistemology (we may infer) is just the application to the ethical case of the general strictures of his foundationalist epistemology.
But Aristotle arguably espouses an analogous but distinct and irreducible sort of foundationalism with respect to what he calls sciences, or *epistemai*, including ethics. Here the claim is that distinct *epistemai* have a foundationalist structure analogous but not reducible to the regress argument at *APo* I.3: each *episteme* has its own distinct foundation.\(^\text{14}\) Thus at *Nicomachean Ethics* I.1-2 (hereafter *NE*), Aristotle offers a regress argument to the effect that human flourishing or *eudaimonia* is the ultimate aim of human action. And so *eudaimonia* is the foundation of ethics. Obviously, it is plausible to think that Aristotle would affirm the conjunction of his own epistemology and his own ‘meta-ethics.’ And yet there is no suggestion in Aristotle’s text, however, that the meta-ethical argument of *NE* I.1-2 establishing *eudaimonia* as the foundation of ethics is reducible to, or an application of, the strictures of Aristotle’s epistemology (i.e., of *Posterior Analytics* I.3). The epistemology and meta-ethics are, to all textual appearances, separate and distinct issues, the connections between them left entirely open.\(^\text{15}\) But while Aristotle himself offers no explicit account of the relation between his epistemology, moral epistemology, and meta-ethics, nothing would prevent subsequent commentators or scholars from doing just that. The product would be (not Aristotle’s but) an *Aristotelian* version of a comprehensive normative foundationalism.\(^\text{16}\)

We could in principle imagine Thomist, Hobbesian, Kantian, Millian, or Moorean versions of this pattern as well. Each of the preceding philosophers is, arguably, an epistemic foundationalist who ratifies the implications of epistemic foundationalism for justified moral belief. But each is also a meta-ethical foundationalist who holds that a justified ethical theory involves a double-tiered structure of foundational and superstructural norms. And yet none of these philosophers clearly or explicitly addresses
the question of how his epistemology relates to his meta-ethics (though Moore perhaps comes closest). As a result, they are in principle committed to the hybrid view under discussion here: they are *simultaneously* committed to epistemic reductionism and to analogy.

It is, of course, possible that the conjunction of reductionism and analogy is redundant. In other words, if epistemic foundationalism is the case, it entails the application of foundationalist strictures to justified moral belief; meanwhile, the supposedly analogous-but-autonomous strictures just mentioned might simply end up being reducible to epistemic norms. That is certainly possible, and if so, would have to be argued for one way or the other. In any case, I take myself to have established the conceptual possibility of the analogy-reduction hybrid, and thus of comprehensive foundationalism. In later chapters, I’ll develop this possibility in ways that turn the mere possibility into a theory worth accepting.\(^\text{17}\)

1.6 Notes


Though Timmons and Hare partly agree with one another, these writers all differ substantially from one another, and from the view defended in this chapter. I have not attempted to offer a detailed discussion of any of the preceding views whether in their own right or in relation to my own, preferring instead to let the contrasts between the enumerated views and my own come out in the course of the discussion.

Perhaps something similar might be said about the varieties of, say, realism, anti-realism, internalism, and externalism. There is little discussion about how all of the varieties of realism—perceptual, semantic, causal, metaphysical, moral, scientific—do or don’t relate to one another. (For partial exceptions to this rule, see Geoffrey Sayre-McCord, “The Many Moral Realisms,” in Essays on Moral Realism, ed. Geoffrey Sayre-McCord [Ithaca, NY: Cornell University Press, 1988], pp. 1-26.) There is even less discussion about how the varieties of internalism and externalism—epistemic, semantic, ethical—do or don’t relate to each other.

Of course, from the perspective of those who deny the existence of a relation between these doctrines, the absence of discussion will seem to be the natural and obvious default position: why expect there to be a literature on, say, semantic and ethical internalism if the two things bear no significant relation to one another? It’s a fair enough question—so long as we assume that they can’t bear a significant relation to one another. But obviously, the sheer absence of a literature on the relation between them can’t count as evidence for the absence of a relation; the presence or absence of a relation requires an argument either way. And we might still wonder: why is it that semantic and ethical internalism look as though they were species of a common genus when they have nothing at all to do with one another?

Thanks to David Benfield, Pierre LeMorvan, and Consuelo Preti for helpful discussions on this issue.


It might be argued that what I’m calling an “analogy” is in fact just a metaphorical extension, but I would argue that metaphors are simply elliptically expressed analogies. For a contrary view, see Donald Davidson, “What Metaphors

4 Cf. the standard definitions of “analogous structures” (or traits), “convergence,” “homoplasy,” and “parallel evolution” in any standard textbook on evolutionary biology. See, e.g., the following definition of “convergence”: “The increasing resemblance over time of distinct evolutionary lineages, in or perhaps several phenotypic respects, increasing their phenetic similarity but generally without associated genetic convergence. Usually interpreted as indicating similar selection pressures in operation. Structures coming to resemble one another in this way are analogous” (both emphases in original), in M. Abercrombie et al., The Penguin Dictionary of Biology, 8th ed. (New York: Penguin, 1992), p. 135.


6 This has the unexpected implication that analogy turns out to be the basis for (or at the very least a sufficient condition of) a sort of externalism about moral motivation.

7 Audi is a foundationalist about justified belief and justified moral belief, but it’s not clear to me that he sees the former commitment as the basis for the latter. They seem to be two separate commitments. See Audi, The Structure of Justification, chs. 1-3, and Audi, The Architecture of Reason, chs. 1-3.

8 The addition of Rawls to this list may seem misleading, as Rawls explicitly rejects intuitionism in A Theory of Justice as elsewhere. I would respond that Rawls’s rejection is itself misleading. Rawls redefines intuitionism in such a way as to make it a much narrower theory than it has been traditionally conceived. Having redefined it in this way, he (oddly) contrasts it with perfectionism. Since perfectionism is not principally an epistemic view, it follows that Rawls must not have an epistemic contrast in mind, and must not be conceiving intuitionism principally in epistemic terms. (Indeed, he explicitly says that he isn’t.) If, however, we focus on intuitionism as an epistemic doctrine, Rawls is—as e.g., Feinberg, Hare, and others have suggested—an intuitionist. He is also an anti-foundationalist (see John Rawls, Justice as Fairness, ed. Erin Kelly [Cambridge, MA: Harvard University Press, 2001], pp. 31-32) and (via his commitment to reflective equilibrium) a coherentist. So Rawls does qualify for inclusion on a list of intuitionists, and represents a paradigm case of the apparent inconsistency I note in the text.


For an exception to this rule, see W. D. Hudson, *Ethical Intuitionism* (New York: St. Martin’s, 1976), who treats intuitionism as an implication of a general commitment to epistemic foundationalism. Thomas Nagel *seems* to be making a similar sort of claim in Thomas Nagel, *The View from Nowhere* (Oxford: Oxford University Press, 1984), pp. 139-49, esp. p. 143.


The Firth-Chisholm debate is, I think, somewhat confused by Chisholm’s contribution to the debate. Firth ascribed reductionism to Chisholm but Firth himself rejected it in favor of something like the view I’ve described as analogy. Chisholm, in turn, accepted the ascription to him of reductionism (Roderick Chisholm, “Firth and the Ethics of Belief,” *Philosophy and Phenomenological Research*, vol. 51, no. 1, [March 1991], p. 119), but ultimately defended a view that was not consistently reductionist (Chisholm, *Theory of Knowledge*, 3rd ed. [Englewood Cliffs, NJ: Prentice Hall, 1989], pp. 58-60). To make matters worse, as a foundationalist, one would have thought Chisholm committed to epistemic reductionism, but Chisholm’s view on the relation between his general commitment to epistemic foundationalism and his moral epistemology is quite obscure. Nor is it clear what consequences the adverbial theory of perception *could* have for moral epistemology, except as the basis for a somewhat eccentric version of intuitionism.


16 The discussion of Aristotle in the text is only illustrative; it would, obviously, take a much longer discussion to do justice to the issues dealt with there. I thank Alasdair Maclntyre for pressing me to clarify my views on Aristotle, though it’s doubtful I’ve done so to his satisfaction.

17 Thanks to Carrie-Ann Biondi for extensive editorial feedback, and to Michael Young for very helpful substantive criticisms on this chapter.
CHAPTER 2:
THE EPISTEMIC APPROACH TO THE FOUNDATIONS OF ETHICS

2.1 Introduction

In one sense, epistemic foundationalism provides an obvious, if not the most obvious, approach to the foundations of ethics. Epistemic foundationalism holds that justified belief has a two-tiered structure, so that beliefs are, if justified, either foundational or superstructural. Suppose that this thesis is true, and suppose that we do indeed have justified moral beliefs. In that case, foundationalism implies that our justified moral beliefs are either foundational or superstructural. So, we might infer, the task of giving an account of the foundations of ethics is simply the application of foundationalist epistemic strictures to the case of specifically moral belief—no more and no less. Either justified moral beliefs are foundational or they’re superstructural or perhaps both. The question is where they go and why they go there.

As I’ve suggested in the preceding chapters, this argument, though correct as far as it goes, oversimplifies things twice over. For one thing, there is more to the foundations of ethics than is captured by the epistemic approach; unless we accept epistemic reductionism, there’s the possibility of an autonomous meta-ethical approach. But even if we confine ourselves to the “application of epistemic foundationalist
strictures to moral belief,” there is more complexity in that thought than may at first be apparent.

As far as the current literature is concerned—and arguably, most of the literature from Sidgwick’s *Methods of Ethics* on—the application of epistemic foundationalist strictures to moral belief takes us directly to some form of intuitionism.¹ For some, this is a reason to reject foundationalism: intuitionism cannot work. For others, it’s a reason to endorse intuitionism: it’s the only foundationalist game in town. For some, it’s irrelevant to the issue of whether one accepts or rejects intuitionism: intuitionism can be endorsed (or rejected) independently of a general commitment to epistemic foundationalism. But few in the contemporary literature seem to have wondered about the inference from foundationalism to intuitionism itself. Is it obvious that foundationalism entails intuitionism? Is it obvious that intuitionism lacks foundationalist competitors in moral epistemology?

I want, in this chapter, to contest the ‘foundationalism-to-intuitionism’ inference, which, as I see it, radically oversimplifies both the varieties of epistemic foundationalism, as well as what it takes to “apply” foundationalist strictures to the domain of moral belief. My aim here is to recast the relationship of epistemic foundationalism to moral belief so as to avoid the reflexive inference to intuitionism, and in so doing, to carve out logical space for non-intuitionist foundationalist moral epistemologies, among them an Aristotelian empiricism based on direct realism about sensory perception.
2.2 Epistemic Foundationalism: Back to Basics

In one sense, the argument for and structure of epistemic foundationalism is so familiar to everyone that it seems an exercise in pedantry to rehearse it again for the nth time. As everyone knows, foundationalism is one of several possible responses to the so-called regress problem, which it solves by positing a definite and non-arbitrary terminus to the justificatory chain. In another sense, however, this apparent familiarity is deceptive, and is, in my view, the reason for a number of hasty inferences and generalizations about foundationalism, among them the intuitionist inference I wish to contest. Almost every contemporary textbook in epistemology opens with an abbreviated account of the regress problem, often presented as though the presentation were itself a formality that the author was obliged to rehearse so as to get down to real business.² But this method of presentation conceals some significant complexity. We can grasp the complexity in question by considering two sets of issues that arise in discussions of the regress, both bibliographical and conceptual.

Two bibliographical oddities arise in these discussions. The first is the pro forma assertion that “the” regress problem was “first” stated by Aristotle in the *Posterior Analytics*.³ There is, of course, a sense in which this is perfectly correct: Aristotle does state an epistemic regress problem at *Posterior Analytics* I.3, and that problem bears a definite similarity to the regress problem as discussed in contemporary epistemology.

But while Aristotle’s regress problem resembles the contemporary one, the differences are equally striking. At face value, Aristotle’s statement of the problem differs from the contemporary one, as does his vocabulary, his presuppositions, and his solution. So does his conception of knowledge, including moral knowledge.⁴ On this
latter point, the interesting fact is that while Aristotle is often described as an intuitionist about moral epistemology, this description has a serious element of anachronism about it. It is questionable whether Aristotle was an intuitionist, and if he was, he was certainly a very different sort of intuitionist from, say, Sidgwick, Moore, Prichard, or Ross.⁵

At a minimum, all of this implies that ahistorical talk of “the” regress problem is a bit of a misnomer. (Historicist talk about there being irreducibly different regress problems is another version of the same misnomer.) If Aristotle’s regress problem is an ancestor of the contemporary version, but both are equally instances of the regress problem, it should follow that both are equally foundationalist. But it should also be clear from reflection on the same facts that there is considerable variation in the forms that foundationalism can take, variety not accounted for in contemporary discussion. One way to see this is to try to translate Aristotle’s Posterior Analytics from the Greek without loss into the jargon of contemporary analytic philosophy. Another and perhaps more poignant way of seeing it is to try to translate a paragraph of contemporary analytic writing on foundationalism without loss into Attic Greek. It’s not easy.

The second bibliographical oddity concerns twentieth and twenty-first century discussion. We are often told, in an equally pro forma spirit, that contemporary foundationalism—i.e., twentieth-century foundationalism—begins with Russell, continues with C. I. Lewis, and is then carried on in the post-war period by Chisholm and Alston, followed by more contemporary authors, e.g., Moser, Triplett, Fumerton, and the like. Indeed, a recent book title suggests that foundationalism is enjoying a bit of a Renaissance after years of neglect.⁶
One minor oddity here is that some philosophers (e.g., Russell and Lewis) are described as foundationalists despite the fact that neither used the term, and neither explicitly described their epistemologies as solutions to the regress problem. A second oddity is that no mention is made of precisely those moral epistemologists whom the ‘foundationalism-to-intuitionism’ inference would lead us to expect to be discussed. After all, if foundationalism underwrites intuitionism, we would expect all intuitionists to be foundationalists, and expect them to be included among the canonical foundationalists. Yet this is only sporadically the case. Russell, Moore, Prichard, and Audi are all foundationalists about justified belief quite generally and intuitionists about moral epistemology, but Sidgwick, Ross, Carritt, Urmson, Kekes, Williams, and McNaughton are intuitionists about moral belief without discernible commitments to epistemic foundationalism. Meanwhile, McDowell, Brody, Nussbaum, Dancy, and McMahan are intuitionists about moral epistemology but anti-foundationalists about justified belief. And Rawls’s authoritative account of intuitionism makes no reference whatsoever to epistemic foundationalism as a motivation for the view.7

The biggest oddity, however, concerns the move from Chisholm as the paradigm foundationalist to the foundationalist Renaissance of recent years. This is not, as it would seem, a “resurrection” of an old-fashioned doctrine, but the transformation of that doctrine into a new doctrine with the same structure. The defining feature of old-fashioned foundationalism was a commitment to self-justified beliefs at the terminus of justification. The innovation of the new foundationalism is a rejection of self-justified beliefs in favor of direct realism about perception, where “perception” is a cognitive but non-doxastic phenomenon.8
These latter issues bring us to the more strictly conceptual ones. A full account of foundationalism involves four distinct tasks carried on at four different levels of abstraction. (1) The first is a meta-epistemic account of the regress problem that is neutral as between the rival solutions to the problem. Here, the question is: what is the problem common to foundationalism and its rivals? (2) The second is a generic foundationalist solution to the regress problem. Here, the question is: what distinguishes foundationalist solutions of the regress (qua foundationalist) from its non-foundationalist rivals? (3) The third is a rigorous defense of a specific form of foundationalism that exhibits the virtues of that particular form as against rival forms. Here the question is: which species of foundationalism best exemplifies the generic foundationalist solution to the regress problem? (4) And the fourth is the application of foundationalist strictures to a particular domain of belief, e.g., scientific, moral, testimonial, etc. Here, the question is: where are beliefs from a given domain (e.g., moral beliefs) located in the two-tiered structure of justified belief as conceived by foundationalism? Are they all in the foundation, or all in the superstructure, or do they straddle both? Ideally, we want an exhaustive account of where every type of belief fits into the structure. Notice that task (4) is itself neutral as between generic and specific accounts of foundationalism, i.e., as between tasks (2) and (3). We can ask where, say, the moral beliefs are located in a generic form of foundationalism, or in some specific form. Notice also that task (4) is to be distinguished from the entirely separate question of how the superstructure of justified belief is to be derived from the foundation.

The conflation of the preceding four tasks—and the default in carrying them out—is (in my view) what masks the complexity of the application of foundationalist
strictures to moral epistemology. One problem arises at task (1). Task (1) defines the problem to which foundationalism is a solution. But if the problem is misformulated, the solutions will be misconceived. A second problem arises at tasks (2) and (3). Task (2) defines foundationalism in generic terms as the solution to the problem identified in (1), and task (3) identifies its species. But if our generic account of foundationalism is too narrow—if, for instance, it conflates genus with species of foundationalism—we’ll get an overly constrained account of both genus and species, and thus miss out on the variety of forms that foundationalism can take. Finally, problems arise at task (4). Task (4) identifies how foundationalism, as previously defined, applies to a given domain. Ideally, it gives us an exhaustive account of how foundationalist strictures of every kind apply to justified moral belief. But if our account hasn’t been exhaustive up to tasks (2) and (3), it can’t be exhaustive at (4). In fact, that understates the problem. If the problem is misconceived at (1), it may be misconceived at (2) and (3); if it is, say, too narrowly conceived at (2) and (3), it will be yet more narrowly conceived at (4). And obviously, if tasks (1)-(4) are neither explicitly acknowledged nor undertaken, we can easily conflate the tasks and fail to see the complexity masked in doing so.

The foundationalist-to-intuitionist inference is, in my view, a result of just such a conflation—or series of them. Because philosophers have failed to articulate the presuppositions of the problem to which foundationalism is a solution, they have failed to make clear the way in which foundationalism and its rivals are mutually exclusive solutions to the same problem. They have also subtly constrained the possibly solutions that foundationalism can take. I clarify that issue in section 2.3. Because philosophers have defined the problem so narrowly, they have defined the foundationalist solution in a
correspondingly narrow way—conflating doxastic species of foundationalism with foundationalism as such. I clarify that issue in section 2.4. Because philosophers have failed to see that the application of foundationalist strictures to moral beliefs is a complex task of its own, they have assumed too quickly that the results of the task are obvious, and lead directly to intuitionism. I dispel that belief in section 2.5.

2.3 Task (1): The Meta-Epistemology of Epistemic Justification

Epistemic foundationalism is one of a set of rival theories of epistemic justification that includes coherentism, contextualism, and infinitism. The task of meta-epistemology is to clarify the subject matter common to these theories, and more to the point, to clarify the problem to which each claims to constitute a solution. Ideally (though it’s not entirely clear that this can be done), a meta-epistemology ought to formulate the problem neutrally as between the claimants, so that the merits and demerits of each theory can be assessed relative to the same problem. In the present case, the subject matter is epistemic justification, and the problem might very roughly be formulated as follows: “Assuming that justification is a necessary condition of knowledge, when are our beliefs sufficiently justified as to meet the relevant condition?” This formulation implicitly makes reference to what I take to be the three interrelated issues of meta-epistemology: (a) the definition of knowledge; (b) the generic and implicit contrast between knowledge and non-knowledge as supplied by ‘justification’; and (c) the specific and explicit character of the justificatory relation. For present purposes, I merely mention these issues so as to distinguish the meta-epistemology of foundationalism from the theory proper. Meta-epistemology as such is beyond the scope of this study.
Obviously, since epistemology is the study of the nature of knowledge, we come to the study of epistemology presupposing that we have knowledge. This may seem at odds with the common supposition that the study of skepticism is an essential feature of epistemology, and in a sense it is. We should distinguish here between two kinds of skepticism: global and local. *Global skepticism* is the claim that we have no knowledge in a given domain. *Local skepticism* is the claim that many (but not all) claims to knowledge in a given domain are illusory. I use the term ‘domain’ here loosely to refer to a set of claims (propositions) united by similarity of content. Thus *global perceptual skepticism* is the claim that we utterly lack perceptual knowledge. *Local perceptual skepticism* is the claim that we lack more perceptual knowledge than we thought we did. Likewise for moral knowledge, or knowledge of biology, mathematics, history, epistemology, and so on. If we take ‘knowledge as such’ to denote knowledge of the world as such—knowledge whose domain is the world—then global skepticism is the claim that we have no knowledge whatsoever.

It is, I think, plainly obvious that global skepticism is incompatible with the possibility of epistemology. If we have no knowledge whatsoever, we lack epistemic knowledge, and we lack a subject matter for epistemology. In that case, epistemology would be impossible, and there would be no point to our having an elaborate discussion of its meta-epistemic presuppositions (or anything else—though we wouldn’t know this). Likewise, if we adopt global skepticism about epistemology, epistemology is impossible. We can only proceed with the task of epistemology if we exclude global skepticism about the world as such, *and* global skepticism about epistemology. Other forms of skepticism—global skepticism about things other than the world as such or epistemology,
and local skepticisms of various sorts—are compatible with epistemology, but as far as I can see, need not be regarded as setting its agenda. (This is not to say that they should be regarded with complacency, either.)

Assuming then that we have knowledge, the first task of meta-epistemology is to identify a manageable set of paradigm instances of it. Paradigm instances of knowledge are instances that indisputably qualify as knowledge, and that (qua indisputable) are sufficiently projectible to help us think about the rest of knowledge.

The trick here is to combine indisputability with projectibility. For any putatively paradigm instance of knowledge, it has to be the case that its qualifying as knowledge is neutral as between the contending theories (indisputability), and its being a token instance of knowledge gives us significant information about the type(s) to which it belongs in order to allow us to generalize about those types (projectibility). Thus a theory that begins with knowledge of God as a paradigm is simply a non-starter: belief in God is the paradigm of a disputable belief, and given God’s *sui generis* nature, knowledge of God will be *sui generis* in content, hence totally unlike the rest of knowledge, and so not even minimally projectible. A theory of this sort begs the relevant meta-epistemic questions right from the start.  

A second trick is to combine indisputability and projectibility with *variety*. In a way, this follows to some degree from projectibility itself. We want to be able to generalize about knowledge—hence the need for projectible paradigms. But our generalizations will be skewed unless we ensure that the paradigms are not just projectible, but tokens of a *variety of types*. It is, for instance, not obvious that all paradigm knowledge takes a doxastic form. Direct realists about perception claim that
perceptual knowledge is both paradigmatic and non-doxastic. Certain theories of memory
make similar claims about iconic memory, as do theories of practical knowledge that
regard skills (how-to-do-X, where that knowledge is not explicitly propositional in the
mind of the doer) as paradigm instances of knowledge. Thus a theory that insists that all
knowledge is doxastic (or worse, confines its paradigms to one type of belief) will be
defective if it excludes (paradigm) non-doxastic knowledge by fiat. Of course, for
theoretical purposes, we can’t multiply variety beyond necessity—hence the need for
manageable variety.

With paradigm instances of knowledge in hand, we go on to formulate a
definition of knowledge in which ‘justification’ plays some major (though not obvious)
role. The so-called standard definition of knowledge as “justified true belief” (JTB) is a
problematic but not entirely implausible starting point. The definition labors under three
potential drawbacks. For one thing, though often described as ‘standard’ and ‘classical’,
it is neither: contrary to many textbooks, the definition is not drawn from Plato’s
Theaetetus, but is (to put it somewhat anachronistically) formulated and rejected there;
and contrary to what is often said, it is disputable that the definition was ever
ubiquitously accepted in any one form in twentieth-century philosophy.11 Second, the
definition is famously susceptible of so-called Gettier counterexamples. Third, the
definition implies that knowledge is a species of belief, excluding by fiat the possibility
of non-doxastic knowledge.

But none of these objections is individually or collectively conclusive as an
objection to the JTB definition of knowledge. As for the first one, obviously, the
definition didn’t have to be standard or classical to be insightful. As for the second one,
even if the definition is susceptible, as stated, of Gettier counterexamples, it is possible that the definition can be repaired to deal with them or reconstrued so that it merely states a necessary condition of knowledge. As for the third objection, we can circumvent it by describing the JTB account not as a definition of knowledge *per se*, but as stating a necessary condition of *doxastic* knowledge (or more precisely, a necessary condition of doxastic knowledge that isn’t self-justified, if there is any). Understood in this way, the formulation captures the important insight that justification is a constituent of knowledge.

The claim “justification is a constituent of knowledge” is ambiguous between two subtly different (and often ignored) construals of justification. On one common construal (call it the *propositional interpretation*), it’s the belief that is the primary bearer of justification. Thus, on this first construal, to ask whether *p* is justified is primarily to ask a question about the justifiability of a belief with a certain content, abstracting from the circumstances of the person holding or inferring the belief. When I ask whether *p* is justified, then, I reflect primarily on the content of *p*, and ask whether, given its content, some idealized or average epistemic agent *S* would be justified in believing or inferring it. Since “justification” primarily modifies *p*, not *S*, the inquiry focuses on the content of *p*, not the epistemic predicament of *S*.

On a less common construal (call it the *personalist interpretation*), it is the person who is the primary bearer of justification. On this construal, the question “Is *p* justified?” is really elliptical for the question whether some *S* is justified in believing or inferring that *p*. Thus “justification” modifies *S*, and only derivatively applies to *p*. When I ask
whether S is justified in believing ρ, I focus on S’s epistemic predicament, then try to see
S’s relation to the belief or inference that ρ.

The difference between these two interpretations is subtle and may seem a matter of mere preference or degree, but the choice may well make a substantive difference to the ultimate character of one’s analysis of justification (e.g., with respect to internalism and externalism, though strictly speaking, each is neutral with respect to them). I prefer the personalist interpretation, and use it in the remainder.

Implicit in our assembling a stock of paradigm instances of knowledge, we have a similar stock of instances of justification. Reflection on this stock reveals a basic contrast between those of our true beliefs whose truth is related only adventitiously to the believer’s cognitive aims, and those whose truth is in fact related to what the believer was trying to accomplish in a given cognitive act. A paradigm case in the former category is a random guess that happens to be true (RTG). Consider a few examples:

S correctly guesses that tonight’s winning lottery number is 5876, despite a total ignorance of what number was chosen, or even what kind of number is typically chosen.

S walks up to a complete stranger at a bus stop and correctly guesses the stranger’s name.

S walks into the Advanced Placement Calculus exam, ignorant of calculus, takes the exam by filling in ovals at random, but manages to score in the 85th percentile.

S, a non-physician with no medical training, listens sympathetically to a friend’s complaints about a nagging headache, leaping immediately to the conclusion that the friend has a brain tumor—which turns out to be the case.
While watching a criminal trial on television, $S$ gets a “hunch” that the accused is innocent—having only as much evidence as can be gathered in sporadic bits by watching dramatic scenes from the trial on the nightly news. $S$ cannot explain why he has this belief or how he acquired it, besides describing it as a “hunch,” but then it turns out that $S$ was right: the accused was innocent.

$S$ goes to a (non-cynical) palm reader who sincerely believes that she can infer biographical facts from the lines on a person’s palm. Relying on nothing but palm-based information, she infers $S$’s marital status, profession, social class, psychological conflicts, and life-aspirations.

$S$ comes to believe in the existence of God on no basis but a commitment to Pascal’s Wager. God turns out to exist, He accepts $S$’s belief, and $S$ goes to heaven.

Assuming ex hypothesi that we can distinguish such guesses from the background of non-randomly formed beliefs, we can rule RTGs out as candidates for knowledge or justification. So $S$’s knowledge that $p$ entails (i) that $p$ is true, (ii) $S$ believes $p$, and (iii) $S$’s belief that $p$ is not an RTG. The latter clause states, in negative terms, the rationale for making justification a constituent of the definition of knowledge. $S$’s knowing that $p$ requires that $S$’s belief that $p$ be linked to the truth-conditions for $p$ by a relation (“justifiedness”) that explains why the beliefs of someone like $S$ generally tend to be true when they are true—i.e., why those beliefs, qua true, are the reverse of RTGs. An account of epistemic justification thus becomes an attempt to cash out clause (iii) in a positive way, eliminating the “not” for some positively-characterized general relationship between $S$’s cognitive aims and the truth of $p$. 
If RTGs are the paradigm case of epistemic non-justification—the clear case that helps us pick out the class of unjustified beliefs—“justification” is the implicit standard by which we do so. To know that a belief is unjustified, I need some rough sense of what it is for a belief to be justified. A theory of epistemic justification is in effect a making-explicit of this implicit norm, the explicit version of which replaces (iii) in our analysis of doxastic knowledge. It’s sometimes said that “epistemic justification” is the placeholder for “that which, when added to true belief, makes something an instance of knowledge.”

The account I’ve just offered gives us a slightly different—and I think more informative—way of putting things. Epistemic justification, we might say, is the placeholder for the general property of true beliefs that differentiates them from RTGs while accounting for why beliefs having that property tend generally and non-randomly to be true.

Let’s assume, then, that if S has knowledge that p, S is epistemically justified that p. But what is it that justifies S in believing p? Equivalently, why is S justified in believing p?—taking the ‘why’ to be a question posed from the “epistemic point of view.” These questions, of course, bring us to the issue of the justificatory regress.

Our paradigm instances of justification either have justification on their own, lack justification but don’t need it, or get their justification from something else. If we understand what it is to fall into the first two categories—a big if—we’ve answered our ‘why’ question. In the last case, however, we have to seek for the justification of (S’s belief that) p in something beyond p, and so we begin on a justificatory regress. Metaphorically, regresses take the form of chains or clusters. The trick here is to
characterize each sort of regress without prejudicing the characterization in the direction of a specific answer.

Suppose that the justification for S’s belief that \( p \) takes the form of a *chain* (i.e., equates justification with linear inference). If so, we understand the justificatory basis of \( p \) in terms of single, discrete items by analogy with links in a chain, e.g.,

S believes that \( p \) on the basis of \( q \), S believes that \( q \) on the basis of \( r \)….

“Chain” in this context is really a metaphor for a series of inferences: in other words, if \( p \) is chain-justified by \( q \), and \( q \) is chain-justified by \( r \), then \( q \) can be inferred from \( r \), and \( p \) can be inferred from \( q \); \( \{p, q, r\} \) constitutes a chain-like regress of inferential justification. One complication here is that the items in question can either be doxastic or non-doxastic, so while it’s tempting to construe the values for the variables \( p, q \) and \( r \) as beliefs, this need not be so.

The idea that regresses of justification consist of “chains” is a ubiquitous assumption in the literature, but it sits side-by-side with foundationalist metaphors that are quite different from that of a chain.\(^{16}\) There is characteristically talk of roots and branches, of hooks and anchors, and so on. We might describe this conception of justification as the *cluster conception*. In its most basic and abstract form, the cluster conception may be represented as follows:

S believes that \( p \) on the basis of \( \{q, r, s…\} \),

where the bracketed variables serve simultaneously to justify \( p \), by contrast with the chain conception, where we have a one-one correspondence between justificandum and justificans.
Once we distinguish these two sorts of regress, however, and make some technical points about the strictly logical properties of the justificatory relation (e.g., about transitivity, asymmetry, the nature of circular and infinite justification), I doubt there is much more we can say about justificatory regresses that is fully neutral between the contending theories. Having characterized justification at the meta-epistemic level, we move now to first-order epistemology, arguing for a generic answer to the regress problem.

2.4 Tasks (2) and (3): Epistemic Foundationalism, Genus and Species

The generic argument for foundationalism goes as follows:

1. If $S$ knows that $p$, $S$ is justified in believing $p$.

2. If $S$ is justified in believing $p$, then either (a) $p$ needs no justification, or (b) $p$ is self-justified, or (c) $p$ needs a justification. In cases (a) or (b), the process of justification stops, and we can infer that foundationalism is the case.

3. If $S$ is justified in believing that $p$, and $p$ needs a justification, then a justificatory regress arises for $S$.

4. If a justificatory regress arises for $S$, then either (a) the regress ends at some non-arbitrary and inescapable terminus (foundationalism), or (b) the regress ends at some conventional and escapable terminus (contextualism), or (c) the regress circles back on itself (coherentism), or (d) the regress goes on to infinity (infinitism).

5. No regress of justification can (qua justificatory) end at some conventional or escapable terminus, circle back on itself, or go on to infinity.
6. Hence if a justificatory regress arises for S, the regress ends at some non-arbitrary and inescapable terminus.

7. Assume *ex hypothesi* that S knows that *p*.

8. Then the justificatory regress ends at some non-arbitrary and inescapable terminus, and foundationalism is the case.

Premises (1), (2), and (3) come from the meta-epistemic analysis laid out in section 2.3 above. Premise (1) tells us that justification is a necessary condition of propositional knowledge, while premises (2) and (3) that justification gives rise to a regress.

Premise (4) lays out the possible solutions to the regress. If we take the ‘or’ here to be exclusive, it follows that the four options are exclusive and exhaustive: the exclusive disjunction trivially guarantees that the four options are exclusive, and a literal reading of the conditional guarantees that the options are exhaustive (taken literally, the conditional would be false in any case in which there were a fifth option over and above the four listed ones). The truth of premise (4) turns on two sub-arguments: a meta-epistemic analysis of the structure of foundationalism, contextualism, coherentism, and infinitism, plus a meta-epistemic account of their exclusivity and joint exhaustiveness of the relevant logical space.

Premise (5) is of course the most controversial, and the one that does the most work in the argument. Its truth depends on generic foundationalist arguments contesting the cogency of contextualism, coherentism, and infinitism.18

Premise (6) follows deductively from premises (4) and (5). Premise (7) is an assumption drawn from our meta-epistemic analysis. Premise (8) follows by hypothetical syllogism from premises (1) (6) and (7), plus a definition of foundationalism.
A full defense of foundationalism is beyond my scope here. For present purposes, I want simply to distinguish between the meta-epistemic account of epistemic justification offered in the preceding section, the generic argument for foundationalism above, and the species of foundationalism that instantiate foundationalism. Once we appreciate the distinctions between these three things, we can quickly see even at this very general level why a commitment to epistemic foundationalism need not entail intuitionism, and in fact is flatly inconsistent with some species of foundationalism.

First, note a difference between the generic argument for foundationalism and the meta-epistemic account of justification sketched in the preceding section. The meta-epistemic account tells us what the epistemic problem is, and so is neutral as between foundationalist and non-foundationalist solutions. The generic argument for foundationalism claims to solve the epistemic problem, and in so doing, defines the broad genus of foundationalist theories, differentiating them from non-foundationalist theories. So understood, foundationalism differs from contextualism in insisting on a single universal standard of justification that transcends ‘contextual’ factors. It differs from coherentism in rejecting circular justification. It differs from infinitism in rejecting infinite justification. Finally, the structure of the argument for foundationalism precludes the possibility of hybrids or fifth alternatives. Foundationalism either gets our full epistemic allegiance or none at all.

Having differentiated foundationalist from non-foundationalist theories, a bit of reflection tells us how the generic argument serves in turn to differentiate species of foundationalist theory. The generic argument tells us that every foundationalist theory has (at least) two components: an account of the foundation, and an account of the relation
between the foundation and superstructure. The first account tells us what sort of item ends the regress and why. The second tells us how the foundation asymmetrically supports the superstructure. Thus species of foundationalism will be differentiated, in the first instance, by the divergent accounts they give of these issues. They’ll differ, in other words, by type of foundation and type of foundation-superstructure relation. In what follows, I confine myself to a discussion of the first of these issues.\(^\text{19}\)

For several decades, it was standard practice in analytic epistemology to assume that the foundational level consisted of self-justified (or “incorrigible”) beliefs, and the superstructural level of beliefs whose justification rested asymmetrically on these beliefs. Questions then arose about what qualified as a self-justified belief, about the nature of self-justification (and the cogency of the concept), and how one got from the class of self-justified beliefs to non-self-justified ones. The so-called “myth of the given” thought to characterize foundationalism consisted in beliefs about appearances, where the beliefs—not the appearances—did the cognitive work of sitting at the foundation of justified belief.

The basic assumption of this literature, which stretches essentially from C. I. Lewis’s *Mind and the World-Order* (1929) up until the 1980s (and to some degree up to the present day), is that all justification is doxastic and inferential: beliefs are the only justificanda in epistemology, and beliefs are the only thing qualified to function as justificans. Since beliefs justify one another by inference, inference is the only possible relation between foundation and superstructure. The question is not what other relation might obtain, but what sort of inference obtains. Call this the doxastic conception of foundationalism.
In the previous few decades, however, a “newer” foundationalist view has arisen which rejects the notion of self-justified beliefs in favor of sensory perception on a direct realist interpretation. This is a “new” foundationalism only in the sense that it had fallen out of favor in mainstream twentieth-century epistemology, but in another sense, it represents a reversion to an older pre- or non-Cartesian foundationalism, whose roots lie in the work of Aristotle, Aquinas, Ockham, and Reid.\textsuperscript{20}

On this sort of view, the foundational level consists not of beliefs but of direct perception of things and properties in the world. Sensory perceptions, so understood, are not beliefs in the sense that they involve no conceptual judgments or inference. And yet they are cognitive items, just the same: they are direct awarenesses or identifications of the concrete features of the world that exist apart from us. As cognitive items they can play a justificatory role, but unlike beliefs, the justificatory role they play is not inferential. The role is, instead, as one proponent of the view puts it, \textit{evidential}.\textsuperscript{21} The senses give us epistemically basic evidence that is not in need of justification. Our cognitive faculties then work this evidence up into concepts, and the concepts, once articulated, become the objects of judgment and predication requiring justification. On this view, no beliefs are foundational; all belief is superstructural. Call this the \textit{non-doxastic conception of foundationalism}.$^{22}$

One obvious lesson that follows from this brief account is that any account of foundationalist theories that simply \textit{identifies} foundationalism with doxastic conceptions of foundationalism is in error. To do so is to conflate species of foundationalism with the genus.
A second lesson concerns a structural difference between doxastic and non-doxastic versions of foundationalism. On a doxastic view, all cognitions take one form: they’re all beliefs. Thus, on this view, one form of cognition can be located anywhere in the structure of justified belief. On a non-doxastic view, however, cognitions take two forms, beliefs and non-beliefs, and the latter are by definition regress-stoppers. Thus, on a non-doxastic view, non-beliefs must be at the foundation, but beliefs may either be at the foundation or in the superstructure. Their being in the superstructure is obvious; in that case, they’ll be supported by non-doxastic cognitions. But their being in the foundation is a possibility, too. A non-doxastic view simply says that non-doxastic cognitions do end regresses. It doesn’t say that they’re the only sort of item that does (or the only sort of item that always does). There are thus two distinct types of non-doxastic foundationalism: one that holds that non-doxastic cognitions are the only regress-stoppers, and another that holds that non-doxastic cognitions are one species of regress-stopper alongside others, but that every non-doxastic cognition stops regresses.

An important implication follows for moral epistemology. Any account of the application of foundationalist strictures to justified moral belief that makes the same identification will compound the preceding error. Conflating species with genus of foundationalism at the general level will conceal from view the very possibility of a non-doxastic foundationalist moral epistemology. But if, as we’ve seen, non-doxastic foundationalisms are possible, and we assume that some of us sometimes have epistemically justified moral beliefs, it follows that these non-doxastic foundationalisms must have implications for moral epistemology. These implications will lead us to
intuitionism if and only if non-doxastic foundationalisms do. But as we’ll see in the next section, they don’t.

2.5 Task (4): Applying Foundationalist Strictures to the Moral Case

We so far have in place the following: an account of the epistemic problem that is putatively neutral between foundationalism and its rivals; a generic account of the foundationalist solution to that problem that is neutral as between rival conceptions of foundationalism; and at least a contrast between doxastic and non-doxastic conceptions of foundationalism that identifies two important (exclusive though not exhaustive) species of foundationalism. Suppose now we ask how the preceding applies to moral epistemology. One question is whether we are led directly to intuitionism. The second and surely more fundamental question is, where are we led, and how? What implications does the foregoing have for moral epistemology, and how do we figure them out? To answer these questions, we have to juggle a series of distinct but interrelated considerations, which I call location, form and content.

We’ve already seen that if S’s cognitions are justified in foundationalist terms, they take a two-tiered structure: foundational and superstructural. When it comes to moral epistemology, one question we can ask is a matter of location: where are the moral beliefs located—at the foundations, in the superstructure, or in both places simultaneously?23

We’ve also seen that S’s cognitions can take at least two forms: doxastic or non-doxastic. So a second question we can ask is a matter of the cognitive form that moral cognitions take: are moral cognitions beliefs or are they non-doxastic cognitions? Notice
that given the account in the preceding section, *form partly determines location*: if \( x \) is a non-doxastic cognition, \( x \) must be foundational, but if \( x \) is a belief, \( x \) can be *either* foundational or superstructural.

Finally, though our subject in moral epistemology is the justification of *moral* belief, recall that a foundationalist moral epistemology is simply the application to the moral case of more general foundationalist epistemology that applies to *all* cognition. Justified beliefs, then, can be either moral *or* non-moral. We are, *ex hypothesi*, interested in the location of moral beliefs. But that interest requires us not just to identify where the justified moral beliefs are located by themselves, but also where they’re located vis-à-vis the justified *non-moral* beliefs. For instance, if there are justified moral beliefs at the foundation, do they share the foundation with non-moral beliefs, or do they monopolize it? Again, we are *ex hypothesi* interested in the cognitive form of moral cognitions. But that interest requires us not just to ask what form the moral cognitions take, but whether the moral cognitions are unique in form or share it with non-moral cognitions. For instance, if all moral cognitions are non-doxastic, is this a unique fact about them, or are some of the non-doxastic cognitions non-moral as well? In short, to get an integrated account of a foundationalist moral epistemology, we have to see it explicitly within the context of the general commitment to foundationalism from which it flows. Since the issue here is a matter of situating cognitions with one type of content vis-à-vis those with others, I’ll call this the issue of *content*.

This distinction between location, form, and content suffices by itself to provide a counterexample to the intuitionist inference. Foundationalism tells us that when justified, \( S \)’s cognitive structure takes a two-tiered form. Qua foundationalism, that’s *all* it says—
no more and no less. All that foundationalism requires is that something be foundational in location and something be superstructural. Granted, form, as we’ve seen, partly determines location: non-doxastic cognitions are always located at the foundation, and never anywhere else. But that is the only way in which form, content, or location determines one another. That exception aside, location doesn’t determine content, content doesn’t determine location, form (for the most part) doesn’t determine either content or location, and content and location (for the most part) don’t determine form.

As a moral epistemology, then, all that foundationalism says is that moral items fit somewhere in the two-tiered structure somehow alongside non-moral items that fit somewhere in there somehow, with the proviso that if moral items turn out to be non-doxastic, they’d better be located at the foundations. Thus foundationalism doesn’t tell us where the moral items must go; it doesn’t tell us what form they must take; it doesn’t tell us what content the foundational or superstructural items must take; and when it comes to beliefs, it doesn’t tell us what location or content they must have. Any combination of moral and non-moral items—and foundational and superstructural ones—is compatible with the basic foundationalist insight that all items must fall into a two-tiered structure. Any such combination is still an instance of foundationalism, and will still have its own unique implications for justified moral belief.

Now, intuitionism in moral epistemology is a species of foundationalist moral epistemology with a very distinctive set of commitments with respect to location, form and content. Intuitions are doxastic regress-stoppers with inherently moral content. That means that if \( x \) is an intuition, it must be located at the foundation of the agent’s cognitive structure; it cannot be superstructural. Second, if \( x \) is an intuition, it has propositional
content, and so must be doxastic; it cannot be non-doxastic. Finally, if \( x \) is an intuition, \( x \) is distinctively moral in content.

Given this, it becomes easy to see that there are at least three obvious sorts of counterexample to the intuitionist inference. Consider a species of foundationalism that holds that justified moral beliefs are all located in the superstructure. Such a view is incompatible with the intuitionist thesis that moral intuitions are justified and located at the foundation. Or consider a species of foundationalism according to which regress-stoppers are non-doxastic. Such a view will be incompatible with the intuitionist thesis that moral intuitions are doxastic regress-stoppers. Or, finally, consider a species of foundationalism according to which intuitions, though doxastic and foundational, are none of them moral in content (e.g., that all intuitions are, say, mathematical or perceptual or aesthetic in content). Such a view, obviously, conflicts with the very crux of intuitionism, which implies the existence of specifically moral intuitions.

Now, we could adduce these three counterexamples and let that suffice as a way of blocking the intuitionist inference. But if we reflect on the counterexamples, we see that there is much a deeper problem for the inference than the sheer fact of its being susceptible of counterexamples. In fact, the “much deeper problem” is not just a problem with the inference per se, but with the way in which the relationship between general and moral epistemology has been conceived. Philosophers have, for one thing, generally not thought explicitly about the relations among location, form, and content as dimensions of foundationalist theories. But (I suspect, under the influence of the propositional conception of justification) they also have not generally seen that the three dimensions interact. If \( S \)’s cognitive structure is epistemically justified by foundationalist strictures,
all of his justified cognitions have a specific location, form, and content at once. If we want to know how foundationalist strictures apply, say, to justified moral belief, we have to ask where the moral beliefs are located, what form they take, and how they stand vis-à-vis beliefs or cognitions with non-moral content, generating all the possible iterations along all the relevant dimensions. If we omit any, we get an incomplete picture of the possible forms that foundationalism can take in moral epistemology.

There is yet another complication here involving the distinction between types and tokens. Notice that a token moral belief must be either foundational or superstructural, but moral beliefs as a type may be split between the foundations and the superstructure: some moral beliefs may be foundational, some superstructural. A token belief must be either moral or non-moral, but moral cognitions taken as a type can in principle be split between doxastic and non-doxastic items. We thus have to be careful when locating items in a foundationalist structure to be clear whether we are talking about types or tokens of an item. What is true of, say, moral beliefs as a type will not be true of any given moral belief-token.

The best way of identifying the sum total of conceptual possibilities is to do so mechanically, and then to prune back the mechanically-generated options as considerations of coherence and plausibility require. As it happens, there are at a very bare minimum 21 possibilities here. It would create excessive complication to discuss each and every one of these possibilities. For present purposes, I’ll simply indicate how the 21 basic possibilities can be generated ab initio, and explain why a large number of these possibilities (eleven, to be precise) are non-intuitionist. In section 2.6, I’ll describe and defend my own favored view.
Let’s begin with a summary of what the preceding account of foundationalism tells us. Suppose that we are trying to identify how the generic strictures of epistemic foundationalism apply to moral epistemology. Assume that we are discussing a cognitive subject \( S \), all of whose beliefs—moral and non-moral—are epistemically justified by foundationalist strictures. Thus any proposition of the form “\( S \) is epistemically justified in believing \( p \)” will be true for any value of \( p \). (We can assume *ex hypothesi* that either \( S \) has no epistemically unjustified beliefs or that if he does, they don’t affect his justified beliefs, and can safely be ignored.) Let’s leave open the possibility that a non-doxastic version of foundationalism may be true, in which case we’re obliged to speak not only of \( S \)’s beliefs but also the non-doxastic items in \( S \)’s cognitive set, taking the term “item” to subsume both beliefs and non-doxastic cognitions such as sensory perceptions. Let us further assume that intuitions (whether moral or non-moral) are a species of belief. Let us assume, finally, that \( S \) has beliefs with a wide range of contents, moral and non-moral, all epistemically justified.

In that case, the following will be true:

1. Every token item in \( S \)’s cognitive set is either doxastic or non-doxastic in form but not both at once.
2. Every item in \( S \)’s cognitive set, whether type or token, is either moral or non-moral in content but not both at once.
3. Every token item in \( S \)’s cognitive set, whether identified by form or content or both, must either be foundational or superstructural, but not both at once.
4. Every type of item in \( S \)’s cognitive set, whether identified by form or content or both, can be foundational, or superstructural, or split between the two.

Claim (1) tells us that every token item in \( S \)’s cognitive set is either a belief or not a belief. Thus the categories “doxastic” and “non-doxastic” are mutually exclusive and jointly exhaustive of items with cognitive content of some kind. Notice, however, that we
can in principle categorize either beliefs or non-beliefs by categories of content, e.g., “moral” and “non-moral.” Nothing about (1) implies that all of the tokens of a given content-type must fall into one and the same category of form. There is no a priori reason, in other words, why all moral items must be beliefs or non-beliefs. For all that (1) says, all the moral items could be beliefs or non-beliefs, or split between the two categories. And likewise with non-moral items.

Claim (2) is just the mirror-image of claim (1). Claim (2) tells us that every token item in S’s cognitive set is either moral or non-moral in content. Thus the categories “moral” and “non-moral” are mutually exclusive and jointly exhaustive of items with cognitive content. Notice, again, that we can in principle categorize moral or non-moral items by form, e.g., “doxastic” or “non-doxastic.” Nothing about (2) implies that all of the tokens of a given cognitive form must fall into one and the same category of content. There is no a priori reason, in other words, why, say, all non-doxastic items must be, say, moral in content. For all that (2) says, all of the non-doxastic items in S’s cognitive set could be moral in content, or all could be non-moral, or they could be split between the two categories. And likewise, obviously, with the doxastic items.

Claim (3) tells us that every token item in S’s cognitive set must either be foundational or superstructural. When it comes to tokens, then, “foundational” and “superstructural” are mutually exclusive and jointly exhaustive of location.

Claim (4) introduces a complication for location that arises when we categorize items by content. Claim (3) tells us that every token item, regardless of form or content, is either foundational or superstructural. A token item falls somewhere within the two-tiered structure, but it can only fall into one place per item. But suppose that we think not
of tokens but of content types—moral and non-moral. While each token moral item will fall in one place in the two-tiered structure, there is no reason to assume *a priori* that all of the moral items as a whole will fall into one place. They might all fall into the foundation, they might all fall into the superstructure, or they might be split between the two.

Claim (4) becomes yet more complicated when we think about categorizing items by formal type. Obviously, any token cognitive item must either be a belief or a non-belief. If we consider beliefs as a category, it is possible to imagine that all beliefs might be superstructural, but I don’t think it makes sense to imagine that all beliefs might be foundational. If all beliefs are foundational, it follows that none are superstructural, in which case we have a one-tiered structure of cognitive items as opposed to a two-tiered one—and so, a departure from foundationalism. Similarly, if we consider non-beliefs as a type, it is possible to imagine that all non-beliefs are foundational, but it makes no sense to imagine that *any* of them are superstructural. I introduced the category of “non-belief” in the first place to denote a non-doxastic regress-stopper. Though it is perhaps conceivable to think of non-beliefs as being located in the superstructure of a subject’s cognitive set, doing so is so obviously incompatible with the motivation for devising the category that it seems better to ignore it.

This is, already, an exceedingly complicated structure. To ease our complications, I propose that we think exclusively in terms of types and ignore tokens altogether. That means that we can completely ignore (3), which is *about* tokens. And in thinking about (1), (2), and (4), it’s likewise best to ignore the clauses that make reference to tokens.
The reason for ignoring tokens is two-fold. First, it is simply too complicated to think about types and tokens simultaneously. Hence it’s best to ignore *something*. The question then becomes: what to ignore? The answer is tokens, not types, which takes us to the second aspect. The relation between claims about types and claims about tokens is after all asymmetrical: what is true of a type is true of every token subsumed by the type, but not vice versa. Thus if we think in terms of types, we can without much complication make the relevant inferences for the tokens within it. But since we can’t do the reverse, focusing on types has the fairly obvious merit of easing our complications while telling us everything we need to know. The distinction between tokens and types is important only insofar as it is possible to make it, and so, possible to conflate the two things. Once we know that the distinction is there to be made, we can safely ignore it and move on.

With all of that in mind, we can generate the basic set of 21 possibilities as follows. Start with the relation between *location* and *content*. Location is a matter of foundation versus superstructure. Content is a matter of an item’s being moral or non-moral. Foundations, we know, can consist entirely of moral items, entirely of non-moral items, or a mix of the two. Call these F1, F2, and F3, respectively:

- **F1**: All foundational items are moral items.
- **F2**: All foundational items are non-moral items.
- **F3**: Foundational items are split between moral and non-moral items.

Superstructures, likewise, can consist entirely of moral items, entirely of non-moral items, or a mix of the two. Call these S1, S2, and S3, respectively:

- **S1**: All superstructural items are moral items.
- **S2**: All superstructural items are non-moral items.
S3: Superstructural items are split between moral and non-moral items.

Combining these two sets, we get nine purely nominal possibilities: (F1, S1), (F1, S2), (F1, S3), (F2, S1), (F2, S2), (F3, S1), (F3, S2), (F3, S3). From this set of nine, two possibilities can immediately be subtracted as incompatible with our initial assumptions, viz., (F1, S1) and (F2, S2). That leaves us with a basic set of seven location-content possibilities.

Now turn to the relation between location and cognitive form. Location, again, is a matter of foundations versus superstructure. Form is a matter of an item’s being doxastic or non-doxastic. Foundations can consist entirely of doxastic items, entirely of non-doxastic items, or a mixture of the two. Call these F4, F5, and F6.

F4: All foundational items are doxastic.

F5: All foundational items are non-doxastic.

F6: Foundational items are split between doxastic and non-doxastic items.

Superstructures, however, can (per our discussion above) only be doxastic; they can’t contain any non-doxastic items. So we have only one possibility here; call it S4.

S4: All superstructural items are doxastic.

Combining these two sets, we get three location-form possibilities, none of which can be eliminated as either incoherent or incompatible with our starting assumptions: (F4, S4), (F5, S4), (F6, S4).

Now, since S’s cognitive set simultaneously combines location, form, and content, we generate our basic set of possibilities by multiplying the first set of seven location-content combinations with the second set of three location-form combinations. That yields a basic set of 21 combinations. Simply by inspection, we can infer that any of
these combinations containing either an F2 or an F5 (or obviously, both) is incompatible with intuitionism. Any F2 combination has a foundation lacking moral items—\textit{a fortiori} lacks a \textit{foundation of moral intuitions}.\textsuperscript{27} Any F5 combination has a foundation lacking doxastic items—\textit{a fortiori} lacks \textit{intuitions altogether} (taking intuitions by definition to be a species of doxastic regress-stopper).\textsuperscript{28} As it happens, there turn out to be eleven such counterexamples to intuitionism.\textsuperscript{29} Interestingly, there turn out to be ten \textit{different} varieties of intuitionism.\textsuperscript{30}

It would be an interesting exercise to work through the internal structure of all eleven varieties of non-intuitionist foundationalism and all ten forms of intuitionism, noting the varieties of each, and noting similarities and differences between them. For present purposes, however, it’s sufficient to note the sheer fact of variety, as well as the fact that the variety furnishes counterexamples to the intuitionist inference.

In the next section, I want to focus on one non-intuitionist combination, viz., (F2, S3, F5, S4). This is a foundationalist epistemology with the following features: (i) none of its foundational items is moral; (ii) its superstructural items are either moral or non-moral; (iii) none of its foundational items is a belief; (iv) all of its superstructural items are beliefs; and (v) \textit{a fortiori}, all of its moral items are beliefs in the superstructure. Obviously, elements (i), (iii), and (v) preclude this view from being a species of intuitionism. I call it \textit{Aristotelian empiricism}, and sketch it in the next section.
2.6 Aristotelian Empiricism as a Foundationalist Moral Epistemology

I take the preceding section to block the intuitionist inference and to establish the possibility of several non-intuitionist foundationalist moral epistemologies. In this section I want, very briefly, to sketch the main lines of my own favored one of these views, namely, Aristotelian empiricism.

On the Aristotelian empiricist view, the foundation of justified belief consists not of beliefs but of the evidence of the senses, understood non-doxastically and regarded as self-evident, and hence as neither justified nor unjustified, but as epistemically basic and not in need of justification. Thus the evidence of the senses confers epistemic justification without needing it; beliefs needing justification—i.e., all beliefs—get their justification, in some complex way, from perceptual evidence. Moral claims, on this view, are all superstructural, and ultimately get their epistemic justification from the evidence of the senses. In terms of our recurring architectural metaphor, justified belief is a building constituted by a (deep and wide) basement of non-doxastic perceptions supporting a huge edifice of beliefs, where the relation between basement and building is evidential, not inferential. Meanwhile, moral beliefs are one floor in the lattice-work of this building, bearing inferential relations to non-moral beliefs. Notice that moral beliefs may, on this view, bear an indirect relation to the evidence of the senses, mediated by “floors” of beliefs more closely related to the “basement” of sensory perception.

This description of Aristotelian empiricism perhaps indicates the obvious problem that the view faces at the level of moral epistemology (whatever other problems it may face elsewhere). Let’s grant for present purposes that the foundation of justified belief consists, as this view holds, of non-doxastic sensory perceptions providing evidence for
the structure of justified belief. At this point, two issues arise, corresponding to what I 
previously called tasks (3) and (4). The first issue is, how in general does the foundation 
relate to the superstructure? How, in other words, do non-doxastic perceptions serve to 
justify belief as such? The second issue is, granting an answer to the preceding questions, 
how does the foundation relate to the ethical domain of the superstructure? How does the 
evidence of the senses serve to justify ethical beliefs? In the space at my disposal, I aim 
just to sketch answers to these questions, and offer what I take to be favorable contrasts 
with rival views in moral epistemology.

The answer to the first question is provided by a theory of abstraction. On the 
Aristotelian view, the evidence of the senses is basic and non-doxastic. It gives us direct 
cognitive access to the objects of the senses, i.e., three-dimensional physical objects and 
their sensory properties. It’s crucial to stress that cognition so conceived is non-doxastic; 
it provides evidence for judgments, but does not itself take the form of judgments. 
Judgment takes the form of concepts, and concepts involve abstraction from the 
perceptual level. “Abstraction,” as one proponent of the view puts it, “is the power of 
selective focus and treatment; it is the power to separate mentally and make cognitive use 
of an aspect of reality that cannot exist separately.”

What we “separate mentally and make cognitive use of” is similarity. Perception 
gives us direct cognitive access to a world of particulars, and qua particulars, each differs 
from all of the others. But some particulars differ more from one another than they do 
from others; particulars differ by varying degrees of differences. Particulars resemble one 
another when they differ from one another less than they differ from others. Abstraction 
is the process by which we pick out salient differences in degree of difference—blues
versus yellows, tables versus chairs, people versus inanimate objects—and retain the
isolated differences as units of cognition, or concepts. On an empiricist view, then, the
epistemic justification of belief is principally a matter of decomposing the belief into its
constituent concepts, tracing the concept back to its perceptual basis, and using the
evidence of the senses as a check on the judgment expressed by the belief.

The answer to the second question is an application of a theory of abstraction to
the special case of ethical concepts. A crucial preliminary to this task is an account of the
similarity that obtains between all ethical concepts—what is the common feature of all
ethical concepts? At this point, of course, one needs a specifically meta-ethical theory
explaining this. A teleological theory will identify some goal as basic, and conceptualize
ethical concepts—i.e., regard them all as similar—in terms of this goal. A deontological
theory will identify some basic duty and conceptualize ethical concepts in terms of this.
This by itself suggests that foundationalist epistemic strictures are by themselves
insufficient to justify ethical beliefs. We cannot, in short, hope to justify ethical beliefs if
all we know is that perception provides the foundation of justified belief, and ethical
beliefs lie somewhere in the superstructure of a foundational two-tiered structure. To get
the epistemic justification of ethical beliefs off the ground we need some theory, itself
epistemically justified, of what all ethical concepts have in common. We need, in short, a
meta-ethics. It follows that an ethical conception that lacks a meta-ethics will not, on
foundationalist strictures, succeed in justifying ethical beliefs. Nor will a theory succeed
in this task that tells us, at the meta-ethical level, that ethical claims are irreducibly plural.
If ethical beliefs are irreducibly plural—i.e., incommensurable—they have nothing in
common. But if they have nothing in common on a foundationalist view, there is nothing for abstraction to do with them, and no way to justify them.

There is thus nothing more to be said about the epistemic justification of ethical beliefs here unless we indicate what they have in common. Not having discussed meta-ethics, I’m not in a position to do that. I turn, then, in the next chapter to the discussion of the meta-ethical component of foundationalism, and return to the epistemic issues in Chapter 6.  

2.7 Notes


2 For a sample, see the texts cited in note 1 of Chapter 1 above.


The distinctions here are meant to highlight different aspects of cross-cutting categories, not to be exclusive.

I take this to be a problematic feature of Alvin Plantinga’s epistemology, which is structured around the assumption that belief in God is “properly basic.” See, e.g.,


17 See references cited in note 15 above.

18 Haack’s arguments against coherentism seem to me to be conclusive (see Haack, Evidence and Inquiry, ch. 3). Though formulated in a context unrelated to
epistemology (the cosmological argument), in my view, William Lane Craig’s Aristotelian arguments against the existence of an actual infinity could, *mutatis mutandis*, be deployed as conclusive arguments against epistemic infinitism (see William Lane Craig, *The Kalam Cosmological Argument* [Eugene, OR: Wipf and Stock Publishers, 1979], pp. 69-102). Though I know of no decisive argument against contextualism, part of the difficulty is that there is no canonical statement of it, either (or at least none that doesn’t collapse into coherentism).

19 I touched on the foundation-superstructure relation in the preceding section’s account of chain versus cluster conceptions of justification, but important as it is for our topic, it would take me too far afield to discuss it in the present context. For good (but very programmatic) discussions, see David Kelley, *The Evidence of the Senses: A Realist Theory of Perception* (Baton Rouge: Louisiana State University Press, 1986), ch. 5, and Susan Haack, *Evidence and Inquiry*, ch. 5.


22 Michael DePaul points out to me that my distinction between doxastic and non-doxastic foundationalisms is ambiguous, non-exhaustive, and/or misleading, for there are intuitionistic versions of foundationalism that involve intuitions that are neither non-doxastic perceptions of the particular nor beliefs. For a view of this sort, see Michael Huemer, *Skepticism and the Veil of Perception* (Lanham, MD: Rowman and Littlefield, 2001), pp. 71-81, and Michael Huemer, *Ethical Intuitionism* (New York: Palgrave Macmillan, 2006), pp. 99-105.

As I’ve drawn it, the distinction between doxastic and non-doxastic foundationalisms is a distinction between foundationalisms whose foundations consist of items with propositional content and foundationalisms whose foundations consist of items with non-propositional content. The controversial enthymematic premise here is that a cognitive item with propositional content is either a belief or so much more similar to a belief than to a non-propositional perception that it deserves to be grouped under the heading ‘doxastic’. This distinction is neither ambiguous, non-exhaustive, nor (I think) misleading. For a sustained defense of it, see Kelley, *The Evidence of the Senses*, chs. 5-6.

Given this, Huemer’s view is—despite his description of it as a form of “direct realism”—a doxastic foundationalism. As he explicitly puts it, on his view, perceptual experience is “propositional” (Huemer, *Skepticism*, p. 74).

In one sense, this is a purely terminological dispute, and nothing much turns on it. Given my way of drawing the distinction, Huemer’s view ends up on the “doxastic” side
of my taxonomy alongside other doxastic views, but in any case firmly distinguished
from the view I defend. Given Huemer’s way of drawing the distinction, my view ends
up being one species of direct realist views alongside his own. He explicitly (though
cursorily) considers and rejects the sort of direct realism I defend in this chapter
(Huemer, Skepticism, pp. 72 and 91 nn. 39-41), and by implication rejects the way I’ve
drawn the doxastic/non-doxastic distinction. As Huemer puts it, if we are perceiving a
bent stick, “the question is whether the senses are telling us that the stick is bent or
whether they are simply giving us certain qualia that we interpret (or are inclined to
interpret) as being caused by a bent stick” (Huemer, Skepticism, p. 71). From my
perspective, however, neither of these questions is right. When we perceive a bent stick,
the senses neither “tell us” that the stick is bent, nor simply convey qualia that we
subsequently go on to interpret. The senses give us a direct apprehension of the stick, but
tell us nothing; we judge the stick bent and predicate bentness of the stick, not of qualia.
Thus Huemer’s way of putting things excludes my brand of direct realism right from the
start.

We could, I suppose, have an argument about whose taxonomy is better. The sticking
point here concerns Huemer’s stipulation that directly perceptual experiences have
propositional but (partly) non-conceptual content (Huemer, Skepticism, p. 74): “[t]heir
content is propositional but (at least part of it is) nonconceptual.” Perceptual content is
propositional, Huemer says, because perceptual experience is representational; it is
nonconceptual because its content does not depend on concepts that the subject has
(Huemer, Skepticism, p. 74). Thus Huemer insists that perceptual experience consists of
“apprehension[s]” that are “assertive mental representation[s],” but are, despite that, not
conceptual (Huemer, Skepticism, p. 79). By contrast, I reserve the term “perception” for
cognitive items that are both non-propositional and (qua non-propositional) non-
conceptual, and so take beliefs to arise (and the term “doxastic”) to apply “above” this
level. I take it to be a virtue of my view that my principle of division clearly demarcates
the purely perceptual (i.e., non-conceptual, non-propositional, non-representational) from
the non-perceptual (i.e., conceptual and propositional, rejecting the representational),
whereas Huemer’s blurs these two categories.

It likewise unclear to me how an assertive mental representation can be non-
conceptual if the term “assertive” is to be taken literally: how can an assertion be non-
conceptual? Huemer seems to suggest that “assertion” is not to be taken quite literally,
however: the “assertiveness” of perceptual experience consists in its “forcefulness”
(Huemer, Skepticism, pp. 77-78). But this seems to me an equivocation involving the
term “assertiveness.” In one sense, Huemer wants to insist that perceptual experience is
“forceful” because the object of perception is undeniably there in the way the objects of
imagination are not. In another sense, Huemer wants to insist that perceptual experience
is “forceful” because it represents propositional content. But these are very distinct
claims: the former (which I endorse) does not require that perceptual content be
conceptual; the latter (which I reject) does.

In Ethical Intuitionism, pp. 99-100, Huemer argues that intuitions are not beliefs, but
his argument there depends essentially on the claims I’ve just discussed in the foregoing.
I thank Michael Young for extensive discussion on these issues.
23 What I’m calling the “location” issue sounds like but is entirely distinct from—i.e., a mere homonym of—what Frank Jackson has in a different context called “the location problem.” See Frank Jackson, *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (Oxford: Oxford University Press, 2000), pp. 1-5.

24 It would be an interesting further exercise to identify principles in virtue of which items of a certain form and/or content were located where they were in the two-tiered structure, but that would multiply our present complications beyond manageable bounds.

25 The first says that “all foundational items are moral and all superstructural items are moral,” which implies that no items are non-moral, and so contradicts our assumption that S has both moral and non-moral items in his cognitive set. The second says that “all foundational items are non-moral and all superstructural items are non-moral,” which implies that no items are moral, and so contradicts the same assumption.

26 A list of all 21 combinations: (F1, S2, F4, S4), (F1, S2, F5, S4), (F1, S2, F6, S4), (F1, S3, F4, S4), (F1, S3, F5, S4), (F1, S3, F6, S4), (F2, S1, F4, S4), (F2, S1, F5, S4), (F2, S1, F6, S4), (F2, S3, F4, S4), (F2, S3, F5, S4), (F2, S3, F6, S4), (F3, S1, F4, S4), (F3, S1, F5, S4), (F3, S1, F6, S4), (F3, S2, F4, S4), (F3, S2, F5, S4), (F3, S2, F6, S4), (F3, S3, F4, S4), (F3, S3, F5, S4), (F3, S3, F6, S4).

27 The F2 combinations: (F2, S1, F4, S4), (F2, S1, F5, S4), (F2, S1, F6, S4), (F2, S3, F4, S4), (F2, S3, F5, S4), (F2, S3, F6, S4).

28 The F5 combinations: (F1, S2, F5, S4), (F1, S3, F5, S4), (F2, S1, F5, S4), (F2, S3, F5, S4), (F3, S1, F5, S4), (F3, S2, F5, S4), (F3, S3, F5, S4).

29 The intuitionist combinations include all of the combinations remaining from the list of 21 after subtracting the F2 and F5 combinations.

I offer a more fine-grained taxonomy of these intuitionist combinations in an earlier draft of this chapter, omitted here for reasons of space. Generally, I think it’s worth distinguishing between versions of intuitionism in which moral intuitions exist only at the foundations from versions in which they exist both at the foundations and the superstructure. It’s also worth distinguishing between versions of intuitionism in which moral intuitions epistemically justify non-moral items and those in which they don’t. For discussion, see Audi, “Intuitionism, Pluralism, and the Foundations of Ethics,” cited in note 7 above.

30 As it happens, in working through the possibilities, one sees that as currently stated, each of the F3 and F6 combinations conceals a slight (but in the present context immaterial) ambiguity which, if disambiguated, gives rise to an additional 18 possibilities
(over and above the original 21), some intuitionist, and some not. Thus it seems possible to generate a total of 39 possible combinations.


32 Thanks to Michael Young and Jason Raibley for helpful comments on a distant ancestor of this chapter, as well as to Alasdair MacIntyre to helpful comments on a (different) ancestor of it. Thanks also to David Keyt, Pierre LeMorvan, and Consuelo Preti for offhand comments that (unbeknownst to them) vastly improved the final product. Finally, thanks to Carrie-Ann Biondi for helpful substantive and editorial feedback.
CHAPTER 3:

THE META-ETHICAL APPROACH TO THE FOUNDATIONS OF ETHICS

3.1 Introduction

In the Introduction of this study, I pointed out that there were two *prima facie* distinct approaches to the foundations of ethics, one epistemic, the other meta-ethical, and in Chapter 1, I raised questions about the relations between them. In Chapter 2, I laid out the logic and presuppositions of the *epistemic approach*, arguing for the possibility of a non-intuitionist foundationalist epistemology for justified ethical belief. In this chapter, I lay out the meta-ethical approach to the foundations of ethics, treating it at least *ex hypothesi* as an autonomous approach to the foundations of ethics.

The task here differs markedly from that involved in describing the epistemic approach. In a sense, the two tasks are mirror-images of one another. In the epistemic case, we began with a general theory, epistemic foundationalism, and applied it to the specific case of justified moral belief. In that case, the existence of a general theory was something we could more or less take for granted; what needed explication was the complex application of that theory to the particular case. In the meta-ethical case, however, the reverse is true. When we survey the range of theories generally accredited in the literature as “foundationalist” what we see are particular approaches to the foundations of ethics that bear what seems at best a family resemblance to one another, but little more.
There is, in other words, a vague way in which Plato’s defense of the Form of the Good in the *Republic* resembles Aristotle’s defense of *eudaimonia* as the *arche* of *politike* in the *Ethics* and Aquinas’s defense of salvation as the final end of human action in the *Summa Theologicae*. Those three accounts, in turn, bear some resemblance to Hobbes’s defense of the laws of nature in the *Leviathan*, Kant’s defense of the categorical imperative in the *Groundwork*, Mill’s proof of the principle of utility in *Utilitarianism*, Gewirth’s defense of the Principle of Generic Consistency in *Reason and Morality*, and so on. But it’s not clear what the resemblance is, and whatever it is, whether it’s enough of a resemblance to justify their classification under a single category of theories. There are similarities that obtain among theories as well as similarities between the theories on the one hand, and epistemic foundationalism on the other. There are differences along those dimensions as well.

Obviously, the task at hand requires us to disentangle the similarities and differences so as to make some sense of things. But in doing this, we labor under all of the difficulties that attend any attempt to categorize projects that resemble each other without explicitly being part of a common endeavor. Aquinas and Mill may have thought of themselves as continuing Aristotle’s project, but it’s not clear Aristotle would have agreed, and it’s doubtful Aquinas would have agreed with Mill (or vice versa) about whose project constituted the genuine continuation. Conversely, Hobbes saw himself as improving on and replacing Aristotle’s and Aquinas’s project (or projects), not continuing it (or them). Meanwhile, Kant saw himself as improving on and replacing all of the projects preceding his, as did Moore. In describing these philosophers as though they endorsed the same theory—“meta-ethical foundationalism”—one runs the risk of
oversimplifying what they said in the name of expository convenience. Of course, in
refusing to see the similarities between their claims simply because the claimants
disavowed such similarities, we run the equal and opposite risk of theoretical myopia.

My aim in this chapter is to describe meta-ethical foundationalism so as to show
what all instances of it have in common as well as how they differ. This deceptively
simple aim is partly precautionary, partly expository, and partly taxonomic. The
precautions are there to forestall confusions in thinking about the issues (section 3.2). The
exposition lays out the basic idea (section 3.3). The taxonomic discussion allows us to
apply the basic idea to issues of greater theoretical complexity (section 3.3).

3.2 Meta-Ethical Foundationalism: Some Preliminary Caveats

As a first approximation, meta-ethical foundationalism is what we might call a
structural analogue of epistemic foundationalism. For epistemic foundationalism, the
subject is epistemic justification, the problem is the epistemic regress, and the solution is
a two-tiered justificatory structure. For meta-ethical foundationalism, the subject is
ethical justification (i.e., right action), the problem is an ethical regress analogous to the
epistemic one, and the solution is a two-tiered structure likewise analogous to the
epistemic one. This apparently simple formulation conceals two potential
misunderstandings that are worth clarifying right from the start. A first and fundamental
caveat requires us briefly to revisit the issue of reducibility. A second issue concerns
terminology.
3.2.1 Reducibility Revisited

In the Introduction of this study, I pointed out that the relation between epistemic and meta-ethical approaches to the foundations of ethics broadly speaking take three forms—homonymy, analogy, and reductionism. (For present purposes, I ignore the analogy-reductionist hybrid.) Homonymy and analogy entail that the epistemic and meta-ethical approaches are distinct, i.e., two ways of describing two different approaches. Reduction entails that they are identical, i.e., two ways of describing one and the same approach. It may be tempting, given some of what I’ve said so far, to infer that I’ve already decided these issues one way or the other, but in fact I intend the exposition in this chapter (and the two subsequent ones) to be neutral on the relation (which I’ll explicitly discuss in chapter 6).

In Chapter 2 (section 2.6), I argued that an Aristotelian empiricist epistemology needs a meta-ethics: “foundationalist epistemic strictures are by themselves insufficient to justify ethical beliefs. . . To get the epistemic justification of ethical beliefs off the ground we need some theory, itself epistemically justified, of what all ethical concepts have in common. We need, in short, a meta-ethics.” Taken by itself, this claim implies that epistemic reductionism is false. After all, if a foundationalist moral epistemology requires a meta-ethics whose claims are formulated independently of that epistemology, the meta-ethics cannot be a mere deductive consequence of the epistemology, from which it follows that epistemic reductionism is false.

Bear in mind, however, that the preceding anti-reductionist conclusion is a consequence not of a commitment to foundationalist moral epistemology as such, but to the Aristotelian brand of foundationalism I outlined in section 2.6. In this chapter, I
abstract from my commitment to that brand of foundationalism, and leave open the
question of whether epistemic reductionism is true or false.

By reverse token: in the paragraph that heads this section, I’ve described meta-
ethical foundationalism as an “analogue” of epistemic foundationalism. This formulation
may seem to imply that I’ve decided things in favor of analogy and against reductionism.
After all, if reductionism is true, it would make no sense to treat meta-ethical
foundationalism as an analogue of epistemic foundationalism, since the one would be
reducible to (not merely analogous to) the other.

But again, I don’t mean to be suggesting at this point that reductionism is true any
more than that it’s false. I merely mean to be suggesting that there is an analogy between
epistemic and meta-ethical foundationalisms, not that the analogy precludes
reductionism.

My aim in this chapter is simply to describe what meta-ethical foundationalism is.
That can in principle be done while prescinding from the analogy vs. reductionism issue,
and there is a clear expository advantage to doing so. We would need to know what meta-
ethical foundationalism is before deciding whether “it” is homonymous with, or
analogous or reducible to epistemic foundationalism. And the only way to do this is to
assume, at least ex hypothesi, that there is something to describe that is prima facie
distinct from epistemic foundationalism. This might turn out to be wrong, but even if
meta-ethical foundationalism were ultimately reducible to epistemic foundationalism, it
would take some involved argument to make the reduction: reducibility is not obvious
from cursory reflection on the content of the two theories, and so, we can reasonably
infer, reductionism is not an obvious expository starting-place.
3.2.2 Choice of Terminology

Having made the preceding point about reductionism, I come now to a related point about terminology. Let’s assume, as per the preceding, that there are two *prima facie* distinct approaches to the foundations of ethics, one epistemic, the other meta-ethical. I’ve *stipulatively* denominated this latter sort of foundationalism “meta-ethical foundationalism,” but this term is somewhat misleading, and the choice of terminology turns out to be far from trivial. As a terminological matter, what we want to capture is the idea of a sort of foundationalism that (a) is about right action, and (b) is structurally similar to epistemic foundationalism, while being (c) *prima facie* distinct from epistemic foundationalism, but (d) possibly being reducible to it, possibly not being reducible to it, and possibly being something to which epistemic foundationalism is reduced.

Almost *any* name for such a view will be misleading. Suppose that meta-ethical foundationalism turns out to be reducible to epistemic foundationalism. In that case, any *distinct* name will be misleading, for in that case, a separate name implies the existence of two distinct things when reductionism shows that there was one thing all along. Even apart from the issue of reducibility, there is going to be a complex story to be told about the relationship between epistemic and ethical normativity, e.g., between the ethics of belief and the ethics of non-doctrastic action. Given this, any *sharp* distinction between the epistemic and the meta-ethical will be misleading if it implies that epistemic and ethical norms are utterly unrelated to each other. This explains why “prescriptive foundationalism” and “normative foundationalism” are misleading: neither prescriptivity nor normativity are unique to ethics. Given all that, it might be tempting to use the name “ethical” or “moral foundationalism” for what I’m calling meta-ethical foundationalism,
but that won’t do, either: after all, the application of epistemic strictures to justified moral belief yields something that could just as easily be called “moral foundationalism,” and the whole point is to distinguish the epistemic from the non-epistemic approaches to the foundations of ethics.

Robert Audi has suggested the term “axiological foundationalism.” But this is problematic, too. For one thing, using “axiological” to mean “ethically choiceworthy” seems misleadingly to imply that doxastic actions can’t be choiceworthy. Second, “axiological” is often used to refer to a sort of value that is incompatible with deontology, but what we need is a term that is neutral as between deontic and non-deontic theories with a foundationalist structure.

“Meta-ethical foundationalism” is not, I admit, a perfect term either, but it seems the least-bad option, and so having issued the preceding caveats, I’ll use the term, understanding it in the stipulative sense implied by the caveats.

3.3 Meta-Ethical Foundationalism: The Common Structure

As with epistemic foundationalisms, meta-ethical foundationalisms both resemble and differ from one another. In this section, I want to identify the common structure of meta-ethical foundationalism, i.e., what it is that all meta-ethical foundationalisms have in common, regardless of any of the other (abundant) differences between them. The account of the common structure should tell us what the structure of Plato’s ethics has in common with that of, say, Pareto, Kant’s ethics with Mill’s, Aristotle’s with Hobbes’s, Herbert Spencer’s with St. Thomas Aquinas’s, etc.
The common structure can, I think, be described as follows. If we look at the paradigm instances of meta-ethical foundationalism, what we find is that, at a certain level of abstraction, they ask the same sort of question about ethics and offer answers with the same sort of structure.

The question in each case is a request for a global justification of moral norms leading to a regress analogous to that in epistemology. This request for global justification turns out to be some version of the “Why be moral?” question. And so each theory has, at least implicitly, an account or interpretation of the problematic that it sees as central to the justification of morality.

The answer to the “Why be moral?” question is, in form, a theory with a two-tiered structure of norms, foundational and superstructural. The foundation of the theory offers a direct answer to the “Why be moral?” question. So any meta-ethical foundationalism will need a defensible, non-arbitrary account of why its preferred foundation really is foundational, i.e., really answers the relevant question. The foundation, in answering the “Why be moral?” question, provides the ultimate unifying reason for why we should be moral, and by implication provides the justification and content of the norms endorsed by the theory. Notice that the task of answering the “Why be moral?” question has an obvious epistemic analogue. Just as, when it comes to epistemic justification, we generate a regress by successive iterations of the question why S believes p, so when it comes to ethical justification, we can in principle generate a similar regress by successive iterations of the question why S ought to □ (or why □-ing would be right for S).
From there, a meta-ethical foundationalism needs a derivation of the superstructure of norms atop the foundation, where the derivation is a hierarchical ordering of principles that applies the content of the foundation to particular cases.

Meta-ethical foundationalism on this interpretation is both comprehensive and (minimally) realist (or “cognitivist”) in its aspirations. As comprehensive, it aims to unify normative claims, aspiring to account for all of them in one integrated account, i.e., one account that makes a systematic unity of the (apparently) disparate norms that constitute morality. Ideally, a foundationalist account of morality leaves no ethical norm, judgment, or claim unaccounted for; every claim has a derivation from the foundation. As realist theories, meta-ethical foundationalisms assert that their normative claims embody moral knowledge, and are by implication true.7

This latter point highlights the importance of a proper formulation of the question that motivates the theory in question. In epistemology, our implicit question is formulated in such a way as to guarantee some connection between foundationalism and truth. If the task is to capture S’s justification for believing that p (i.e., for believing that p is true), to trace S’s belief to the foundation of justified belief is ipso facto to trace it to the basic items in S’s cognitive make-up, where the “basic items” are basic in virtue of being S’s basic form of access to reality. If these items really are basic, and really do access the world, and S’s beliefs are relevantly based on them, S’s beliefs are justified in virtue of their relation to the world, and so have a chance of being true.

On a meta-ethical foundationalist view, the guiding question is some version of “Why be moral?” The assumption here is that if that question is legitimate and properly formulated (on some specific interpretation of it), then it has an answer, and the answer to
it must be bivalent. Since “the answer” is the foundation of ethics, the item or items in the foundation play a role analogous to the basic items in a foundationalist epistemology. In epistemic foundationalism, basic cognitive items constitute the knower’s basic form of access to reality. In meta-ethical foundationalism, the foundation picks out the most ontologically basic truth-makers of moral propositions. So if the answer to the question is true, and genuinely responds to the question, the foundational claim is true, and provides the ultimate true answer to the question of why we should be moral. Assuming that the superstructure of norms is properly based on the foundation, its claims will be true, as well.

I’ve emphasized here that unlike epistemic foundationalism, meta-ethical foundationalism has not necessarily been self-consciously endorsed under that (or any) description. In the twentieth and twenty-first centuries, at least, epistemic foundationalists have typically operated under the assumption that, whatever their disagreements, they had something in common with other foundationalists. But this hasn’t usually been true of meta-ethical foundationalists.

For that reason, when we take the “common structure” of meta-ethical foundationalism and apply it to real-world instances of the theory, we should expect the picture to be somewhat messy, yielding stronger and weaker versions of commitment to the theory. It might help to have a conceptual apparatus in place to deal with this complexity.

Many theorists who qualify as meta-ethical foundationalists do so implicitly as opposed to by stated intention, and so instantiate meta-ethical foundationalism to varying degrees. Thus some theories are only implicitly foundationalist: they happen to instantiate
meta-ethical foundationalist strictures without ever announcing an intention to do so. Within this category, there will be *degrees of approximation in (implicitly) instantiating those strictures*. Thus a theory can, without explicit intention, approximate foundationalist strictures nearly perfectly, or else do so very imperfectly. One theory may offer a foundation without making explicit the question that motivates it. Another theory may offer a foundation without deriving a superstructure from it. And so on. By reverse token, some philosophers have *explicitly aspired to meta-ethical foundationalism, but fallen short* of that aspiration in various ways and to varying degrees. So a theorist might claim to have posed the relevant sort of question and offered a two-tiered answer without being successful in the attempt.

Among strong meta-ethical foundationalisms, I would count those of Aquinas, Hobbes, Spinoza, Kant, Mill, and some of their more recent followers. In each of these cases, I would argue, we have instances of an *explicit* commitment to all three elements of meta-ethical foundationalism—a global request for justification, and a two-tiered theoretical structure answering that request, with a foundation doing so explicitly and a superstructure extending the answer to particular cases. Likewise, in each of the cases, I would argue, there is—given the sophistication and thoroughness of argument—a high likelihood of success at the foundationalist aspiration. Among weaker foundationalisms, I would count Plato and Aristotle. In these cases, we arguably have a sort of proto-foundationalism—an inexplicit and implicit instantiation of foundationalist strictures without a self-conscious commitment to them.
3.4 Differentiating Meta-Ethical Foundationalisms: The Erotetics of Ethics

What I’ve so far had to say about the common structure of meta-ethical foundationalisms is, necessarily, very abstract and generalized. All such theories pose a similar sort of question, and offer a similar sort of answer. An account of the common structure is, of course, a prerequisite of discussing the differences between versions of the theory: a discussion of sheer differences between meta-ethical foundationalisms would not be a discussion of versions of one theory unless we knew what common features united the different instances. But with the common structure in hand, it becomes apparent that the differences between the theories take on a new significance: they become differences along shared dimensions. Thus, all meta-ethical foundationalisms pose a similar sort of question, but each question differs, sometimes radically, from the others. Likewise, all meta-ethical foundationalisms offer a similar sort of answer, but the answers differ radically as well.

The question “Why should I be moral?” was first asked—and rejected—in that precise form in English-speaking philosophy by F. H. Bradley in his 1879 work, *Ethical Studies*. The point has to be put in that cumbersome way because the Question is not literally one question but one of a family of related questions, a fact that collides with philosophers’ tendency to assimilate all of the different questions into one.

What is common to all instances of the Question is a request for a global justification of morality, along with a motivation to fit this justification. What’s striking, however, is how different these requests end up looking. Some take the conflict or apparent conflict between morality and self-interest to be basic, and ask how the conflict can be reconciled or explained away. Some take the teleological character of action to be
basic, and equate the justification of morality with the task of identifying some ultimate end of human action. Others take rationality to be basic and ask how moral norms instantiate it. Some philosophers emphasize the justificatory task, de-emphasizing the motivational one: to know why we ought to be moral is not to be motivated to be moral. Some philosophers do the reverse, and some seek to do both simultaneously. No aspect of the problem is incontestable, and most formulations of the guiding question are either left implicit or, when made explicit, are vague and equivocal. Further, the content of a given philosopher’s question is affected by the non-ethical background assumptions within which it operates, e.g., theism versus atheism, libertarianism versus compatibilism, and the like.

As the questions differ, so too do the answers. Though we could in principle differentiate meta-ethical foundationalisms by type of question or by type of answer, it turns out to be easier to do the latter: theorists have been more explicit about the answers they have wanted to give than the questions they have wanted answered, and in any case, certain questions lead naturally to corresponding answers.

At the broadest level, I think we can distinguish non-naturalist from naturalist foundationalisms. Non-naturalist theories locate the foundations of ethics outside of the natural realm—on their own criteria for distinguishing between the natural and the non-natural. We might, under this heading, distinguish theistic from non-theistic versions of non-naturalism. Theistic foundationalisms hold that God’s sovereignty and omnibenevolence provide the foundations of ethics. On this sort of view, our ultimate reason for being moral makes essential reference to a deity with these traits: we should be moral because God’s sovereignty and omnibenevolence demand it. One clear version of
such a theory is a divine command ethics, according to which moral claims are true and
overriding because God wills them to be so.\textsuperscript{11} Another version of such a theory might be
a theistic teleology according to which moral claims are true and overriding because God
has given the world an objective normative structure that demands compliance and
rewards its practitioners for compliance.\textsuperscript{12}

\textit{Non-theistic versions of non-naturalism} identify the foundations of ethics with a
non-natural dimension that is not a deity. Moore’s conception of the good as a non-
natural property in \textit{Principia Ethica} is a paradigm example of this view; Plato’s
conception of the Form of the Good at \textit{Republic} VI probably qualifies as well.\textsuperscript{13} On views
of this sort, we should be moral because, given their inherent character, certain (non-
natural) axiological properties merit our attention and allegiance in a normatively basic
way.

Naturalist theories locate the foundation of ethics within the natural realm, again
on their own criteria. We may subdivide naturalist theories into three categories—
impersonalist, contractarian, and egoist.

\textit{Impersonalist} theories hold that the deliverances of reason, understood as agent-
and temporally neutral prescriptions of a human faculty, provide the foundation of ethics.
Our ultimate reason for being moral is an ultimate impersonal prescription that binds us
independently of our goals, needs, or desires, and applies in an invariant and categorical
way to all members of the moral community. I count two basic subcategories of
impersonalism.\textsuperscript{14} \textit{Deontological} theories (e.g., Kant’s) identify a basic duty, regarding
that duty as morally basic, and defining all further norms in terms of it.\textsuperscript{15} \textit{Intrinsic value

theories (e.g., utilitarianism) define certain states of affairs as valuable in themselves, and define right action as the maximization of the value of those states of affairs.¹⁶

Contractarian foundationalisms (e.g., Hobbes’s) hold that our ultimate reason for being moral derives from a social act of consent or agreement, whether hypothetical or actual, implicit or explicit. We should be moral, on this view, because we have agreed to be so, and our own agreements are ultimately what bind us to act. The foundation of this view consists of an account of the promise that binds us to morality; the superstructure consists of the prescriptive implications of this promise.¹⁷

Finally, egoistic foundationalisms hold that each agent ought to be the ultimate intended beneficiary of his or her own actions. For the egoist, then, we should be moral because our overall self-interest depends on it. An egoist theory, then, will consist of an account of why our interests are the basis of practical norms, what our interests consist of, and how the general injunction to follow our self-interest leads in turn to particular moral norms.¹⁸

3.5 Conclusion: The Outstanding Issues

My aim in this chapter has been principally expository and taxonomic—to lay out the common structure of meta-ethical foundationalism, capturing what all instances of the theory have in common, while distinguishing it from the epistemic approach, and distinguishing each version of the theory from other versions. That leaves us with two outstanding issues, one pertaining to meta-ethical foundationalism as such, the other pertaining to its relationship to epistemic foundationalism.
Some opposition to foundationalism about ethics is opposition to the epistemic approach. In other words, it rests on opposition to epistemic foundationalism. But some opposition is detachable from the epistemic issues and pertains to meta-ethical foundationalism. With the framework before us, we can isolate three types of objection.

Erotetic objections are the broadest sort of objection to meta-ethical foundationalism: they contest the legitimacy of the “Why be moral?” question, and thereby question the legitimacy of the foundationalist enterprise as such. Foundational objections accept the legitimacy of the enterprise, but contest the adequacy of a given answer (or any answer). Finally, casuistic questions accept the legitimacy of question and answer, but contest the adequacy (or possibility) of a given derivation (or any derivation) from the foundation. In Chapters 4-5, I take up the most common and broadly ranging erotic objection to meta-ethical foundationalism, namely, the view that the “Why be moral?” question is either illegitimate or incoherent. Having answered this objection, I’ll be able to take for granted that the foundationalist enterprise is a sound one.

Having canvassed both epistemic and meta-ethical foundationalisms, we’ll then need to revisit the issue I raised in the Introduction: how, if at all, do they relate to each other? Are they reducible to one another? If not, how do they fit into one integrated theory? In Chapter 6, I answer this question by sketching the features of a comprehensive normative foundationalism that combines a commitment to a non-intuitionist foundationalist epistemology at the level of epistemic foundationalism and a strong ethical egoism at the meta-ethical level. The result, I suggest, is not the mere conjunction of two unrelated theories, but an integrated and mutually supporting whole while being consistent with epistemic foundationalism.
3.6 Notes


4 Nor has it characteristically been treated that way in the literature, whether classical or contemporary. In fact, the reverse has typically been the case. Aristotle does not tell us that the meta-ethical regress argument of *Nicomachean Ethics* I.1 is reducible to the epistemic regress argument of *Posterior Analytics* I.3. Nor does Kant tell us that the argument for the categorical imperative in the *Groundwork* is a mere deductive consequence of a commitment to epistemic foundationalism; indeed, most commentators tell us that he is not an epistemic foundationalist at all. In more recent philosophy, neo-Kantians like Hare, Gewirth and Korsgaard offer meta-ethical foundationalisms while explicitly rejecting epistemic foundationalism—as do utilitarians like Brink, Aristotelians like MacIntyre, contractarians like Gauthier and Narveson, and moral realists of the Cornell realist variety.


6 This formulation is deliberately ambiguous about whether the “ought” in question is overriding or not.


13 Moore, *Principia Ethica*, chs. 1 and 6; Plato, *Republic*, 505a-511e. Moore’s stipulation that the good is a simple, undefinable and non-natural property puts him in this category (see *Principia*, secs. 5-14), as does Plato’s claim in the *Republic* that the Form of the Good is “superior in rank and power to being” (509b) along with his general ontology of Forms in the middle dialogues.


16 I’ve deliberately avoided describing this group of theories as “consequentialist,” for reasons explained in the Introduction, but I intend it to apply to those consequentialist theories (excluding egoism) which fit the description in the text. Cf. entries for “Consequentialism” and “Utility and the Good” by Philip Petit and Robert E. Goodin, respectively, in *A Companion to Ethics*, chs. 19 and 20.


18 I take Aristotle and Spinoza to be classical exponents of egoism. See especially Aristotle, *Nicomachean Ethics*, IX.4-8 and Spinoza, *Ethics*, IV, Proposition 18. For a
CHAPTER 4:
META-ETHICAL ANTI-FOUNDATIONALISM AND THE EROTETIC DILEMMA

4.1 Introduction

Though meta-ethical foundationalism as I described it in Chapter 3 includes a large variety of theories with a long history and a great deal of structural complexity, it is I think very much the minority view in twentieth and twenty-first century ethics. The majority view is what I’ll give the name “anti-foundationalism,” which should be understood to denote the denial of meta-ethical foundationalism as such, as by analogy atheism is the denial of theism and nominalism the denial of realism.

What bears emphasis here is the point of the modifier “as such.” There are many aspects to and versions of foundationalism, and thus many things about it to deny or dispute, but for present purposes, I reserve the term “anti-foundationalism” for the principled rejection of (meta-ethical) foundationalism.¹ What makes someone an anti-foundationalist is his understanding of what foundationalism is, and his rejection of it for what it is—not merely of some aspect or version of the doctrine, nor even the conjunction of every version of it, but of the principle behind it. As we’ll see, this narrows the idea of anti-foundationalism to a very specific pattern of argument, and allows us to give some coherence to what would otherwise be a sprawling literature of disconnected polemics. In fact, I think it allows us to reduce the whole of that literature to one basic argument.
In section 4.2 of this chapter, I’ll explain in some detail what I mean by a “rejection of foundationalism as such,” with special attention to the meaning of the italicized phrase. In sections 4.3-4.6, I’ll lay out in detail what I take to be the most powerful anti-foundationalist argument so construed, which I call “the erotetic dilemma.”

4.2 Anti-Foundationalism as the Rejection of Foundationalism ‘As Such’

As we’ve seen from Chapter 3, meta-ethical foundationalism is a form of moral realism about right action inspired by an analogue of the regress problem from epistemic foundationalism. On this view, a sound moral theory begins by posing a fundamental question about the justification of right action, provides an answer to that question which functions as the foundation of ethics, and derives a comprehensive structure of norms from that foundation. An anti-foundationalist, as I’m using the term, is someone who rejects “foundationalism as such,” where the “as such” qualifier serves to include criticisms of what is essential to foundationalism and to exclude criticisms of what is accidental to it.

As I see it, the “as such” qualifier serves to exclude three common sorts of argument often misleadingly described as “anti-foundationalist.” The first sort rejects foundationalism on the basis of what it claims to be an affinity between foundationalism and totalitarianism (or terrorism). The second rejects foundationalism by rejecting as illusory the metaphysical and epistemological presuppositions of ethics. The third rejects foundationalism by rejecting paradigm instances of the theory. Whatever the merits of these arguments, I’ll suggest, none really deals with what is central to foundationalism, and so none qualifies as a form of anti-foundationalism in the sense I intend here.
4.2.1 Objection from the Analogy to Totalitarianism

One criticism of foundationalism—more common outside of philosophy than in it, but expressed by philosophers as well—rejects meta-ethical foundationalism on the basis of its supposed affinity with political totalitarianism or terrorism. The argument here is roughly that totalitarian regimes and terrorist groups with totalitarian aspirations have typically relied for their legitimacy on ideologies of some sort, and the ideologies most congenial to totalitarian rule have, however crude, been foundationalist in structure. Such ideologies isolate one value as normatively basic, and seek to construct an ethos based on it that encompasses the whole of human life, enforcing those norms with the coercive power of the state. Since, as I observed in Chapter 3, foundationalist theories do aim at comprehensiveness, there is (this criticism asserts) a strong affinity between foundationalism and totalitarianism. Since totalitarianism is obviously immoral, it follows that foundationalism is unacceptable.

For present purposes, I propose to exclude such criticisms from the scope of “anti-foundationalism,” because whatever their legitimacy as observations about the nature of totalitarianism, they don’t at all count as criticisms of foundationalism. In its strongest form, the “totalitarianism” objection holds that every totalitarianism is foundationalist (i.e., based on some form of foundationalism), but even if this were true, it wouldn’t follow that every foundationalism was totalitarian. At best, the criticism in question alerts us to ways in which foundationalism can be exploited for political purposes, and thus suggests that a foundationalist theory ought to account for ways in which its prescriptions
are particularly liable to political manipulation. But as any view can be exploited for some nefarious political purpose, even at its best the “totalitarianism” criticism tells us nothing about foundationalism as such, and attempts to deploy it that way, though common, are straightforwardly fallacious. So no instance of the analogy-to-totalitarianism argument counts, for my purposes, as an instance of anti-foundationalism.

4.2.2 Objections from the Rejection of the Possibility of Ethics

A second sort of view might attack foundationalism by attacking the very presuppositions of ethics as such. For instance, if global skepticism were the case, we would lack knowledge altogether, and so lack any capacity to produce or defend any sort of ethical view, foundationalist or otherwise. Likewise, were hard determinism true, responsibility would fall by the wayside as an illusion, and so in turn might any conception of morality involving a conception of responsibility. So it is with specifically morality-debunking versions of determinism. There is a family of deterministic theories—e.g., those of Marx, Nietzsche, Freud, possibly Darwin, and their successors—that hold that morality is an epiphenomenon of some deeper deterministic factor to whose operation it is reducible without remainder. On Marx’s view, morality is a delusion of class prejudice; on Nietzsche’s, it is a function of the will to power; on Freud’s, it is a sublimation of repressed sexual urges; and on Darwin’s, it is an expression of naturally selected traits.

Again, whatever the merits of such views, they are too broadly formulated to count as anti-foundationalist: they’re objections to the very possibility of ethics as such. I don’t mean to suggest that there are no arguments to be made against such views. Indeed,
a successful foundationalist theory might go some of the way toward answering the challenges contained in them. But these challenges also depend on claims that are anterior to any specifically moral theory, e.g., skepticism and determinism. The rejection of skepticism and determinism are presuppositions of (meta-ethical) foundationalism; they’re not established by it. So any meta-ethical foundationalist would have to respond to the challenges presented above prior to and independently of a defense of meta-ethical foundationalism. Hence such global challenges to morality as skepticism or determinism fall outside the scope of the dispute between foundationalism and anti-foundationalism.

4.2.3 Objections to Foundationalism by Refutation of Its Instances

A third sort of objection rejects foundationalism by rejecting what it takes to be a paradigm instance (or instances) of foundationalism, taking the paradigm instance to represent the view as such. Thus a critic might select a particular foundationalist theory—say, utilitarianism—identify its foundationalist character, attack utilitarianism, and conclude that foundationalism itself had been rebutted. This is a remarkably common criticism of foundationalism, and though in its crudest forms it seems (and is) obviously fallacious, in its subtler forms, the fallacy is harder to identify.⁶

On the face of it, the argument just described commits what would seem the obvious fallacy of conflating genus with species: even if a theorist successfully refutes utilitarianism (or egoism or whatever), utilitarianism is but one species of the genus of foundationalist theories, and a refutation of one species cannot be said to refute every other species—any more than the extinction of one biological species would mean the extinction of the entire genus (assuming more than one species). Since there are in fact
instances of the fallacy in the literature that are just this crude, we can simply note that this sort of objection is a fallacy, exclude it from anti-foundationalism, and leave the matter there.

But some instances of the argument form are subtler. Philosophers frequently attempt to rebut foundationalism by taking, say, utilitarianism or egoism or Kantianism as the paradigmatic or most powerful forms of foundationalism, and rebutting them. The implicit claim here is that if theory \( x \) is the most defensible form of foundationalism, and it fails, then foundationalism must itself fail. This claim is itself a form of the more general principle that if a theory \( x \) belongs to type \( K \), and \( x \) is the most defensible instance of \( K \), then the refutation of \( x \) is a refutation of \( K \). But plausible as this mode of argument may appear—and ubiquitous as it may be—it is as fallacious as the “crude” version rehearsed above.

A first problem is that the argument turns on \( x \)’s being shown to be the “most defensible” form of foundationalism. This is rarely, if ever, actually shown. Indeed, it is rarely explained what the criteria for “most defensible” are; for that matter, it is rarely even shown that the targeted form of \( x \) is the most defensible version of \( x \). More fundamentally, however, the phrase “the most defensible instance of \( K \)” is really elliptical for “the most defensible instance of \( K \) so far,” and the inference from “most defensible instance of \( K \) so far” to “most defensible instance of \( K \)” is simply a non-sequitur—unless we have a demonstration in hand that no future instance of \( K \) could be more powerful than the most powerful instance of \( K \) so far. Again, this latter inference is rarely, if ever, cashed out.\(^7\)
It’s worth noting that the preceding point would apply even to a successful refutation of the conjunction of every known instance of every foundationalist theory ever devised. To sink one foundationalist theory is to sink that particular theory. To sink every foundationalist theory is still merely to sink each particular foundationalist theory without saying anything—even by implication—about the category “foundationalist.” Here my earlier biological analogy breaks down. The extinction of every species of a biological genus does indeed destroy the genus (and given Dollo’s Law, destroys it forever). But the refutation of every species of theory under a genus does not refute the generic thesis motivating the specific theories—not even if the theories refuted were the most defensible instances of each theory so far formulated in human history.

This latter point should be obvious, but it is almost ubiquitously overlooked in the literature, in ethics and elsewhere. Philosophers often argue against a whole category of theories not by arguing directly against what is genuinely definitory of the category, but by arguing against prominent instances of the category. Since it is rarely feasible to argue against every instance, the arguments usually target the best-known instances, or the most popular instances, or the instances most current in the literature, or simply as many instances as space will allow in a given context. The implicit innuendo behind such arguments is tacitly inductive: “If the best instance of theory \( K \) we’ve so far seen has collapsed, we might as well assume that every instance of \( K \) will collapse in the future.” In the present context, the claim becomes: “if every version of foundationalism we’ve so far seen has collapsed, we might as well assume that every version in the future will collapse.” But even if proponents of such arguments actually demonstrated that the target theories were the “most defensible so far,” they overlook the facts that conceptual novelty
is always possible, and so, that robust theories have an abundance of resources to overcome objections. Given this, it should be obvious that the most defensible instances of foundationalism so far in history are not necessarily the most defensible instances possible, so that the refutation of present instances tells us nothing about the defensibility of future permutations—unless the refutation explicitly targets what is definitory of the theory.

It may seem quixotic to some to suggest that if every instance of a theory in the last 2,500 years has failed, this shows nothing about the theory as such. But it really doesn’t. It may show instead that the attempts to rebut the theory have make the same mistake 2,500 years in a row. Anyway, assume ex hypothesi that humanity is to last another 25,000 years and that our rate of philosophical progress is roughly the same over the next 25,000 years as it’s been over the past 2,500. In that case, all things considered, every instance of theory $K$ so far may not amount to all that many instances.

The lesson would seem to be this. If an objection to a moral theory doesn’t target the specifically foundationalist features of the theory and doesn’t generalize to foundationalism as such, it will not rebut foundationalism simply to pile on instances of objections. On the other hand, if a refutation of a single foundationalist theory does so in a way that not only rebuts that theory but does so by showing what’s common to all foundationalist theories, no further refutation of other instances will be needed.
4.2.4 The Structure of Anti-Foundationalism

To reject foundationalism, then, is to reject what is definitory of it, not merely what instantiates it. We can, I think, distinguish rejections of foundationalism in two ways:

(a) by which element of foundationalism is being rejected.
(b) by the degree to which the element is being rejected.

Consider (a). We’ve seen that a foundationalist theory consists of a question, a demonstration, and a derivation. We could therefore imagine three sorts of rejection of foundationalism. The strongest rejection would reject the very ‘askability’ of the motivating question, and \emph{a fortiori} reject both the cogency of any answer to it, as well as the defensibility of any derivation from it. If the idea of a foundation of ethics gets its very point from the Question, and the prescriptive derivations get their very point from the rationale for the foundation, then the rejection of the Question constitutes a full rejection of the foundationalist project. Call this an \emph{erotetic objection}. A somewhat weaker objection might accept the ‘askability’ of the Question (or of some version of it), but deny that it \emph{could} be answered, and \emph{a fortiori} deny the legitimacy or relevance of any derivation from it. Call this a \emph{foundational objection}. A yet weaker rejection might accept both the ‘askability’ of the Question and the defensibility of some answer or answers to it, but reject the possibility of deriving any substantive normative conclusions from this foundational principle (or, of deriving a sufficiently robust system of norms from it). Such objections might assert that foundationalist derivations fail because foundationalist theories cannot in principle vindicate all or most or some significant number of our background moral beliefs. Call these \emph{casuistic objections}.
Now consider (b). Under each of the preceding three types of objection, we could imagine stronger and weaker versions of each rejection, differentiated by the vehemence or explicitness with which the foundationalist position was rejected. We need this further distinction because anti-foundationalists have varied greatly in both respects vis-à-vis foundationalism. Two philosophers might concur in rejecting all three elements of foundationalism—question, foundation, and derivation—and yet one philosopher might do so vehemently and explicitly, while the other might do so equivocally and implicitly. It’s tempting to overlook the latter altogether, unless we have some category in which to put him. And it’s the stronger versions that will qualify most clearly as anti-foundational.

This taxonomy gives us a total of six possible anti-foundationalist positions, with three basic positions, and two variations on each position:

1. Strong erotetic anti-foundationalism.
2. Weak erotetic anti-foundationalism.
4. Weak derivational anti-foundationalism.
5. Strong casuistic anti-foundationalism.
6. Weak casuistic anti-foundationalism.

Each of these positions has been held by philosophers at some point, but I think it’s clear that the strongest position is (1), which implies (3) and (5)—which are themselves explicit versions of (2) and (4). The strongest case for anti-foundationalism, then, would stem from the strongest arguments for (1): if (1) is true, then (3) and (5) are true, and foundationalism is an utter failure. On the other hand, if (1) can be refuted, foundationalism clears its first hurdle, and has a *prima facie* chance at success. Given this, in understanding anti-foundationalism, it makes the most sense to focus on (1). In the next few sections, I’ll lay out a well-known argument for it in some detail.
4.3 Strong Erotetic Meta-Ethical Anti-Foundationalism

The erotetic objection to meta-ethical foundationalism takes its canonical form in F. H. Bradley’s 1876 essay “Why Should I Be Moral?” Bradley’s thesis is re-stated and elaborated on in H. A. Prichard’s “Does Moral Philosophy Rest on a Mistake?” and then reintroduced into post-war moral philosophy in Stephen Toulmin’s *The Place of Reason in Ethics*, thereafter to become a staple of the literature in English-speaking analytic philosophy. Though I cannot pretend to have done a literally exhaustive search on the subject, it is safe to say that the vast majority of writing on the subject in English-speaking analytic philosophy—Australian, North American, and British—follows the Bradley-Prichard-Toulmin line in rejecting the legitimacy of the “Why be moral?” question and accepting the soundness of the erotetic objection to foundationalism. The argument is structurally simple: Foundationalism is motivated by a question of global scope about the justification of ethics. If the question is senseless, so is the task of answering it. Since foundationalism just is that task, foundationalism fails in a fundamental way: it lacks a coherent motivation, and so, never gets off the ground.

More specifically, the erotetic objection takes the form of a constructive dilemma that I call the Erotetic Dilemma (hereafter, the dilemma). It goes as follows. Take some questioner, $Q$, and start with the assertion that there is such a thing as the distinctively “moral point of view” (to be defined below). To ask a question from within a point of view is to take that point of view as authoritative; to ask it from outside of a point of view is to ask it while consciously rejecting it. Then,
1. If $Q$ asks the Question from within the moral point of view, the Question is *pointless*.
2. If $Q$ asks the question from outside of the moral point of view, the Question is *illegitimate*.
3. $Q$ can only ask the Question from within or outside of the moral point of view (i.e., there is no morally neutral point of view from which the Question can be asked).
4. Hence the Question is either pointless or illegitimate.

From (4), it’s easy to secure the conclusion that the anti-foundationalist wants, and thus to undermine foundationalism. Having undermined the foundationalist project, the anti-foundationalist can help herself to any or all parts of the project that may appeal to her, without committing her to foundationalism as such. The point is that she can do so liberated from the confining constraints of foundationalism; she has no reason to take seriously the (now unmotivated) injunction to answer the Question, or to force her normative theorizing onto the Procrustean methodological bed required by foundationalism. She can also theorize without paying much attention to the metaphysical or epistemological worries that accompany foundationalism, e.g., questions about the truth-conditions for moral propositions, or the sort of epistemology required for moral realism.

On the face of it, the dilemma is obviously valid: it’s a simple case of constructive dilemma. Determining its soundness, however, is a remarkably difficult affair. The difficulties arise from two quarters. One obvious problem concerns the truth of the premises. A less obvious difficulty arises from the equivocality of the terms in each premise. In fact, the former issue turns out to be largely derivative of the latter, since what makes the premises controversial is the fact that the question “Why should I be moral?” and the companion notion of “the moral point of view” are highly contested (or
at least, contestable). Though the contestability of these concepts is often hidden from view in mainstream theorizing, it’s important to bring their complexity fully into view if we’re to assess the strength of the anti-foundationalist case.

4.4 The Case for the Dilemma: Premise (1)

Let’s begin by considering premise (1) of the dilemma, and spelling it out as explicitly as possible.

Premise (1): If \( Q \) asks, “Why should I be moral?” from within the moral point of view, \( Q \)’s asking “Why should I be moral?” is pointless.

Two terminological issues intrude before we can assess the truth of this premise. First, what is a “pointless question”? Second, what is “the moral point of view”?

The claim that “Why should I be moral?” is a pointless question is a staple of the literature. But it’s not always clear what is being claimed when the Question is described as “pointless.” Note, however, that the structure of the dilemma requires that we distinguish two distinct ways in which the Question fails: it’s pointless from within the moral point of view, but illegitimate outside of that point of view. Obviously, then, “pointlessness” and “illegitimacy” have to be distinct deficiencies in the Question.

A full explication of the idea of a “pointless question” would require a long excursion into the fascinating field of erotetic logic.\(^{11}\) For my purposes, however, I’ll offer a relatively informal account. First, we need a definition of question, after which we can specify the notion of a pointless question.

We should begin by distinguishing rhetorical from literal questions, and focus on literal questions, setting aside rhetorical questions as a relatively peripheral form. Central
to a literal question is the idea of a sincere request for information. We can see this by considering the six interrogative particles in English—who, what, when, where, how, and why. It would be tedious to go through each case, but each, I think, can in some way be characterized as involving a request for a specific piece of information, specified by the grammatical form of the particle.  

A subsidiary but important point about questions is that the request for information embodied in a question always involves at least one—and typically many more than one—tacit assertion as a condition of the intelligibility of the request. In other words, in asking something, it’s always the case that the questioner is simultaneously asserting something else, where the assertion provides the background conditions that give sense to the original request for information. “Who” typically presupposes a social context (or at least, a context in which we can pick out different persons); “when” presupposes claims about the flow of time and events; “where” presupposes claims about spatial relations; “how” presupposes background beliefs about causal relations; “what” presupposes a context of discourse in which we have in hand individuation-conditions for the objects being discussed; and “why” presupposes background beliefs about justification and/or explanation. Obviously, since a question is a request for information, one assumption that every question makes is that it’s possible and meaningful to seek information and ask others about it—an obvious fact, but still a tacit one, since the fact itself can’t be asserted in the form of a question.

A pointless question, I want to say, would be a request for information whose tacit conditions of assertion rendered superfluous the activity of requesting the information requested in the first place. The clearest way of doing this would be to ask a question
which requested information that was already obviously included in the tacit assertions that got the question off the ground, e.g., “Do I exist?” or “Is it possible to formulate a question in language?”

Now let me move to “the moral point of view.” This is a little harder and more diffuse a concept; to explain it, we need to proceed in stages. First, we need the concept of a point of view. Then we need an analysis of the commonsense or conventional conception of morality, using "conventional” here in the purely stipulative sense of “conventional as described by the twentieth-century Anglo-American philosophical literature.” This conception of morality, we’ll see, concerns itself primarily with the concept of moral obligation, which in turn, is characterized as impersonal, categorical, and justificatory (in senses to be defined). With this conceptual apparatus in hand, we can see how the notion of “moral” modifies “point of view” to arrive at the distinctive notion of “the moral point of view.”

A point of view is a psychological perspective we take on a distinctive subject-matter, which gives rise to a distinctive set of considerations, or reasons for thought or action. The phrase arises both in technical academic discussions as well as in colloquial speech. In technical contexts, a “point of view” often names an academic discipline: we can speak of something from a scientific, biological, psychological, logical, economic, or historical point of view. Sometimes it names a worldview: a naturalistic versus a religious (or theistic, Christian, Islamic) point of view; realist or nominalist points of view; or as Rawls puts it, the point of view of a representative person behind the veil of ignorance. And sometimes, a point of view names some relatively determinate field of endeavor: military, strategic, political, pedagogical, recreational, aesthetic, and so on. Generally, if I
take something “strictly from an \( x \)-point of view,” I have at least a rough idea of an \( x \)-perspective on some domain of belief or action, as distinguished from non-\( x \) perspectives, and I focus on \( x \)-considerations while excluding to the extent possible, all non-\( x \) considerations. Clearly, the \( x \)-perspective must not only be determinate, but normatively determinate: when I look at the world (or some part of it) “from an \( x \) point of view,” I not only look at it in a specific way, but acknowledge that I have to act a specific way.

The moral point of view is the perspective we take when we resolve to let distinctively moral considerations override contrary considerations, whether ‘practical’ or ‘theoretical’. That, of course, pushes the issue to the definition of ‘moral.’ There is a voluminous literature on the definition of “morality,” but it’s essential to anti-foundationalism that there is one intuitively clear conception of morality, of which specific moralities are mere variations. The basic concept of morality is that of obligation—the idea of something we should, ought, or must do. Though philosophers disagree about the precise overriding force of moral oughts, it’s undeniable that there are at least some contexts in which moral oughts override non-moral oughts, which is to say that there are some contexts in which morality dictates that we ought to do such-and-such, and its dictating this overrides all other considerations. In such contexts, we say that one ought to do such-and-such all things considered. So considered, the moral point of view is a distinctive outlook on the world which literally transforms the way it looks to the person who adopts it. In doing so, one brings out the world’s morally salient features at the expense of its non-moral ones, filtering out the latter, and adopting principles to govern one’s actions—usually in a social context. The moral point of view, as traditionally conceived, has three essential features.
The first feature is the *impersonality* of moral principles on this view. In taking the moral point of view, one explicitly excludes considerations of self-benefit in assessing the moral worth of an action; one sees oneself as merely one member of the moral community, each one of whom is equally and agent-neutrally significant in one’s deliberations. Impersonality so conceived is typically regarded as constitutive of morality as such, and is therefore neutral between conceptions of morality—Kantian, Utilitarian, Contractarian, Theistic. The differences between these mainstream theories are differences within the moral point of view, not alternatives to it. On the standard conception, then, though self-interest is at least occasionally a permissible motive of action, it remains at odds with the distinctively moral motive of action. Though we may make concessions to self-interest, to the extent that we act on a specifically morally motive, we must put self-interest aside. While philosophers in the past few decades have begun to take issue with the stringency of the impersonality constraint, and with some of its particular implications, virtually no philosopher rejects the constraint as such.\(^\text{15}\)

A second feature of the moral point of view is the *categoricity* of moral obligation, as opposed to its conditionality or “hypotheticality.” An obligation is conditional if its binding force is incurred by some voluntary act, and can be relinquished by some voluntary act. The paradigmatic example of a conditional obligation is a promise: one makes a promise by explicitly agreeing to undertake some action. If I promise to ship you a crate of books tomorrow, my promise binds me to the action, but if I’ve made no promise to you, I have no positive obligation to you whatsoever. An obligation is categorical if its binding force exists independently of any voluntary act, and cannot be escaped by any voluntary act. On the standard or mainstream conception of
morality, whatever their superficial grammatical form, moral principles are categorically binding, as opposed to conditionally so: morality consists of an important class of duties that we owe one another without our having incurred them by any act. As Rawls puts it,

[I]t is characteristic of natural duties that they apply to us without regard to our voluntary acts…Thus if the basic structure of society is just…everyone has a natural duty to do his part in the existing scheme. Each is bound to these institutions independent of his voluntary acts, performative or otherwise.  

Likewise Stephen Darwall:

While other elements of Kant’s theory of morals…perhaps reflect rather poorly any very widely shared view about morality, his insistence that moral requirements are categorical imperatives expresses our common sense about an important part of ethics. 

And James Rachels:

Commonsense morality would say, then, that you should give money for famine relief rather than spending it at the movies.

This way of thinking involves a general assumption about our moral duties: it is assumed that we have moral duties to other people—and not merely duties we create, such as by making a promise or incurring a debt. We have natural duties to others simply because they are people who could be helped or harmed by our actions. 

The language of “binding,” as J. L. Mackie and others have pointed out, involves the metaphor of a sort of cord that “ties” the agent inescapably to various duties. The description of this duty as “natural” suggests its inescapability: it is as much “tied” to us as any of our faculties. In another sense, however, the description of such duties as “natural” is highly misleading, since our natural features are amenable of naturalistic explanations. By contrast, most (though not all) philosophers—and all anti-foundationalists—regard the binding force of natural duties as not further subject to
explanation or analysis. This is clearly brought out by the way in which Rawls, Darwall, and Rachels appeal to conventional beliefs to explain the concept: for Rawls, this takes the form of an appeal to reflective equilibrium; for Darwall and Rachels, an appeal to “common sense.”

A third feature, less fundamental than the first two but still important, is morality’s *publicity*. The moral point of view requires that for any action we take, we be prepared to justify that action sincerely to others.

Let’s call the preceding conception of morality morality$_c$, taking the subscript to refer to “conventional” or “commonsense” morality. Our account of morality$_c$ allows us to make our Question more precise as well, since the definition of “morality” governs the meaning of two of the other terms in the Question—“why” and “should.” Philosophers typically distinguish between two sorts of “why”-questions: the justificatory why (why$_j$) and the explanatory why (why$_e$), the first of which seeks reasons, the second of which seeks causes. Similarly, there are (at least) two senses of “should”: the moral should (should$_m$) and the non-moral should (should$_{\sim m}$), the first being the sense of should we take from the moral point of view as described above, the second subsuming all senses of should that fall outside of that perspective.

The preceding characterization of the moral point of view makes it easy to select the version of the Question favored by the anti-foundationalist for premise (1). Premise (1) tells us that $Q$ accepts the moral point of view. The moral point of view is a psychological perspective in which one takes distinctively moral$_c$ reasons as overriding, and we know that moral$_c$ reasons are impersonal and categorical in their binding force and public in character. That fact obviously implies that $Q$ rejects should$_{\sim m}$ for should$_m$. 

134
We’re then left with the choice between why$_j$ and why$_c$. Most philosophers have argued that explanation—in the sense of giving a causal explanation for an action—is not justificatory. Justification is intrinsically reason-based, not causal; causal explanation is “normatively inert.”

That suggests that if $Q$ accepts the moral point of view, $Q$ accepts why$_j$ over why$_c$.

We can now make premise (1) perfectly precise. When $Q$ asks

(Q1) “Why should I be moral?”

she is in fact asking

(Q2) “Why$_j$ should$_m$ I be moral$_c$?”

And premise (1) becomes:

(1*) If $Q$ accepts the authority of the moral point of view, and asks “Why$_j$ should$_m$ I be moral$_c$?”, $Q$’s question is pointless.

In plain English, (1*) ascribes the following extended thought to $Q$: “Given that I accept the authority of conventional morality, what is the rational justification for why I should, in the distinctively moral sense of ‘should’, be conventionally moral?”

Given a specification of pointless questions and the moral point of view, the truth of premise (1*) should be obvious. If $Q$ were to assert the antecedent of (1*), she would in that very assertion have to grasp that her question had already been answered. And so the question is pointless.

Consider the following cases:

Case 1: In asserting the antecedent, $Q$ grasps that her question is already answered. But then the question is pointless, and premise (1) is true.

Case 2: In asserting the antecedent, $Q$ doesn’t quite grasp that her question has already been answered; what has to be shown to her is that, unbeknownst to her, it
has. As stated, the Question remains pointless, and we simply have to repeat the preceding argument until \( Q \) overcomes her mental block in seeing its pointlessness.

*Case 3:* In asserting the antecedent, \( Q \) doesn’t at all grasp that her question has already been answered. This case can be divided into subcases, corresponding to the different reasons why she might not grasp the pointlessness of the Question.

*Subcase 3a:* \( Q \) doesn’t in fact understand what she is saying. But we then explain this to her, to the extent of her amenability and our patience. The Question remains pointless.

*Subcase 3b:* \( Q \) really does seem to understand the terms of the Question, but insists that she doesn’t see its pointlessness. At this point, we’re forced to conclude either that \( Q \)’s mental faculties are defective (or temporarily blocked), that she is being insincere (or dogmatic or stubborn), or that she’s so inept at introspection that her own mental states are opaque to her. Each of these subcases obviates further discussion of the Question; if \( Q \) is in one of these states, she’s incapable of rational thought. Each therefore leaves the consequent of premise (1*) intact and leaves the Question pointless.

I’ve belabored what may seem an obvious point in part because the literature treats it somewhat cavalierly, and in part because outside of the context of academic philosophical discussion, what may seem “obvious” actually becomes quite difficult to see. It’s simple enough to “imagine” an agent who asks the Question “from within the moral point of view,” and easy enough to set up an interpretation of “the moral point of view” so as to make the Question look pointless. But moral reality rarely complies neatly with such schematizations, and so we need to make the point more concrete. To what in moral reality might the preceding discussions actually apply? Cases (1), (2), and (3a) are cases in which \( Q \) asks the Question but can’t see its pointlessness due to conditions beyond her control. Case (3b) is one in which \( Q \) won’t see the pointlessness of the Question. The explanation for both cases is similar, I think. In real life, moral agents do not self-consciously ask “Why should I be moral?” from within the moral point of view.
by attaching subscripts to the terms they use. But they do ask confusing questions without subscripts. Sometimes, a question can be motivated by confusion, other times through corruption of consciousness. In either case, the complexity of the question abets confusion.

4.5 The Case for the Dilemma: Premise (2)

4.5.1 Explaining Premise (2)

Premise (2) says this:

Premise 2: If \( Q \) asks “Why should I be moral?” from outside of the moral point of view, \( Q \)’s asking “Why should I be moral?” is illegitimate.

Since we now know what the moral point of view is, we can easily infer what it is to stand outside of it, and what it to accept or reject it: it’s to reject any one of the fundamental suppositions of morality. But what is an “illegitimate question,” and why is the Question made illegitimate merely because of the perspective from which it’s asked?

A question, as we’ve seen, is a request for information that makes tacit assertions as a condition of the request. An illegitimate question would be one whose tacit assertion-conditions were in some sense “impermissible.” But it’s difficult to pin down this sense of “impermissibility,” especially in a context in which the question posed is about the justificatory status of moral concepts, and is being asked from “outside of the moral point of view.” Even in purely epistemic contexts, the notion of permissibility has a kind of moral flavor, so that explicating the illegitimacy of a question in terms of its impermissibility seems circular or uninformative.
One way of filling out the relevant sense of illegitimacy is to say that an illegitimate question is a question-begging question—a question that tacitly makes assertions that commit the fallacy of begging the question. But of course, this move merely transfers the issue from the question’s “illegitimacy” to the right way of characterizing the fallacy of begging the question. And here we confront a somewhat embarrassing problem: the phrase “begging the question” is used rather loosely both in colloquial speech and in philosophy, and as Douglas Walton has pointed out, the standard textbook treatments of the fallacy are often misleading.22

One suggestion is that a question-begging argument is a circular argument. But doesn’t that beg the question against coherentists, who regard circularity as a virtue? Another suggestion is that the fallacy arises when Smith’s argument that \( p \) depends on some premise \( q \) that Smith’s interlocuter Jones refuses to accept. But an interlocuter’s refusal to accept your premise can’t make an argument fallacious; what if the refusal to accept \( p \) is itself based on fallacious reasoning or otherwise irrational? The fallacy was initially identified and formulated by Aristotle in his *Topics*, *Prior Analytics*, and *Sophistical Refutations*, but Aristotle’s account of the fallacy, precise as it is, is too complicated and controversial for present purposes. In any case, it appeals to the same quasi-moral notion of ‘being entitled’ to assert a proposition in a dialectical context that I mentioned earlier and put to one side.

Since I lack the space here to offer a full theory, it might help simply to work from an example. Take the standard textbook example, “Have you stopped beating your spouse yet?”—asked of a witness in a court of law. The grammar of the question obviously presupposes that the witness once was a spouse-beater. Suppose, however, that
the witness had never beaten his or her spouse. Suppose, in fact, that the witness had never been married, and had never assaulted anyone. But suppose, further, that the witness is required to respond to the question as stated, in an unequivocal manner, and in terms of a “yes” or a “no.”

There are two cases here. In Case 1, the witness answers “yes,” intending a denial.

The witness’s intention is as follows:

Witness’s Intended Response (Case 1): “Yes—I have stopped beating my spouse in the odd sense that having never been married, and having never assaulted anyone, a fortiori, it’s false that I ‘stopped beating my spouse.’”

But the jury might reasonably infer a bare “yes” to mean:

Jury’s Inference (Case 1): “So he once used to beat his spouse, but now he’s stopped.”

In Case 2, the witness answers “no,” intending a denial. The intended response is as follows:

Witness’s Intended Response (Case 2): “No—I haven’t stopped beating my spouse, because I never got started in the first place. In fact I’ve never assaulted anyone, and I’m not even married!”

But the jury might well infer the bare “no” to mean:

Jury’s Inference (Case 2): “So he’s still a spouse-beater!”

*Ex hypothesi,* we know that the witness is innocent of beating. We also know that the witness is trapped by the very terms of the question, and will remain trapped so long as the question is asked and the answer demanded in this way. We might, then, understand an illegitimate or question-begging question as one whose tacit conditions of assertion are *truth-blocking* with respect to some subject-matter. In this case, the question “Have you
stopped beating your spouse?” is truth-blocking with respect to the guilt or innocence of
the witness. A truth-blocking question, then, is one such that if one insists on asking that
question, and getting a directly responsive and unmodified response, there is by the very
nature of the question, no practical way of getting an answer that’s true. The tacit
assertion-conditions of the question interfere with the truth of any possible answer.

Now we can look more carefully at premise (2). \( Q \) rejects the moral point of view
in premise (2). We have two variants of three of the terms of the Question—two variants
of “why,” two of “should, and by default two of “moral. Since \( Q \) is rejecting the moral
point of view in premise (2), we can ipso facto reject the possibility that \( Q \) is asking any
one of the interpretations of the Question in the preceding section. \( Q \) must therefore be
asking one of the following questions:

Explanationist Questions: Why\(_e\) should\(_{m/-m}\) I be moral\(_e\)?
Subordinationist Questions: Why\(_j\) should\(_{m/-m}\) I be moral\(_j\)?

To defend premise (2) of the dilemma, what has to be shown is that each of these
questions is illegitimate in the sense just described. Let me take each question in turn,
explaining what each asks, and evaluating them from an anti-foundationalist perspective.

4.5.2 Explanationist Questions

In plain English, either form of the Explanationist’s Questions might be put as
follows: “What is the causal explanation, involving either the moral or non-moral should,
for why I should be moral in the conventional sense?”\(^{23}\) The Explanationist, in other
words, wants a causal explanation for why we “should” be moral. Now, strictly speaking,
we might have been able to generate four versions of the Explanationist Question, putting
subscripts on “should” (should\(_m\rightarrow m\)) and on “morality” (morality\(_c\rightarrow c\)). But for all practical purposes, there is really only one question here. The essential feature of an Explanationist Question is its request for a causal explanation rather than a non-causal rational justification of ethical norms; it is specifically the appeal to why\(_e\) that puts Explanationist Questions outside of the moral point of view. For ease of exposition, I hold morality\(_c\) constant in considering Explanationist case, treating questions pertaining to non-conventional conceptions of morality (morality\(_c\)) separately.

What about should\(_m\) and should\(_m\)? The very fact that Explanationist Questions seek causal explanations for moral norms makes the difference between these two options superfluous as well. What makes Explanationist Questions reductionist is their attempt to reduce the “should” of “Why should I be moral?” to a causal relation, which is another way of trying to reduce the “should” to an “are.” To understand the Explanationist’s Question, then, we have to move from

“Why should I be moral”

to

“Why should I be moral?” asked from outside of the moral point of view,

to

“Why should I be moral?” asked as a request for a causal explanation, i.e.,

“Why\(_e\) should\(_m\rightarrow m\) I be moral\(_c\)”

which—substituting “are” for “should”—becomes the question,

“What is the causal explanation for the fact that persons are moral, on the conventional conception of morality?”
Some version of this question is at the heart of contemporary approaches to Sociobiology. Taking altruism to be a defining feature of conventional morality, and taking the findings of evolutionary biology to have crucial bearing on morality, the preceding question becomes:

What is the causal explanation for the possibility of altruism in organisms that are the products of natural selection, given that natural selection seems to select against altruism?

The task here is to show that moral properties (e.g., a capacity for altruism) supervene on natural properties (e.g., properties explainable by evolutionary theory) by means of bridge laws.²⁴

The anti-foundationalist response to the Explanationist is that a request for a causal explanation for morality is truth-blocking with respect to the subject, for two reasons. A first reason bears on freedom and responsibility: causal explanations are deterministic, but if moral obligation and responsibility are essential to morality, and both entail metaphysical freedom, and freedom is incompatible with determinism, then a request for a causal explanation for morality is misconceived. This objection, of course, presupposes the truth of metaphysical libertarianism—a controversial claim. But even if we accept a compatibilist interpretation of freedom, a further problem arises: causal explanations miss the normativity of moral norms. Moral value is inherently authoritative; moral norms guide action. Causal explanations, by contrast, are normatively inert; at best, they goad action. In looking for something that goads rather than guides, the Explanationist is by the terms of her question looking for something that
morality is not. And it’s not surprising that when she looks in this way, she doesn’t find what she’s looking for. The Explanationist, then, misses the autonomy of morality.²⁵

It’s worth noting that the anti-foundationalist needn’t prove that morality is autonomous; she can merely point out that if it is, the Explanationist will obscure that truth with his ill-conceived question. If morality is autonomous—as on many accounts it is—then Explanationist Questions at least run a risk of being illegitimate. While this doesn’t demonstrate their illegitimacy, it creates doubt as to it, and pushes the burden of proof onto the Explanationist to justify the legitimacy of the question. The Explanationist can’t merely ask “Why, be moral?” in a vacuum; he needs explicitly to defend his presuppositions before his question can be taken seriously. The anti-foundationalist’s claim will be either that Explanationist’s fail to offer such justifications, or that the justifications they do offer fail.

4.5.3 Subordinationist Questions

Setting aside the Explanationist, Q might be asking a somewhat different question from outside of the moral point of view.

Subordinationist Question: Why, should, be moral,? In English: What are the non-moral reasons for acting morally in the conventional sense? The Subordinationist, in other words, wants to subordinate morality to some non-moral end or consideration, i.e., some motive that conventional morality takes to be non-moral. His request for a justification (as opposed to a causal explanation) for morality, is compatible with the moral point of view. What puts him outside of the moral point of view is the sort of thing he counts as a justification, namely, the making of morality as a
means to some end that is not itself moral by conventional standards. Consider two versions of this question.

The most obvious example of non-moral value is pleasure on an egoistic-hedonistic account of it, or on a preference-satisfaction account of utilitarianism. There are few real examples of hedonism in recent analytic philosophy; one notional example is Sidgwick’s account of it in *The Methods of Ethics*, itself a refined form of Epicureanism. Examples of it abound in popular culture and the social sciences, however, and strains of it are audible in the literature on “moral saints.” Though easy to dismiss out of hand, some versions of it can be surprisingly subtle and sophisticated. On some interpretations, the view also has a respectable lineage that stretches back via Sidgwick to Mandeville, Epicurus, and other Hellenistic thinkers. A second example of a non-moral value is political order or peace. The *locus classicus* is Machiavelli’s statement of the view in *The Prince*; a contemporary version of it is the so-called “dirty hands” argument.

On the anti-foundationalist view, the trouble with the Subordinationist’s request is that a justification of morality in the service of an avowedly non-moral end fails precisely if it is successful. Such a “justification” makes morality a means to something that is not moral. Some of the ends that qualify as “non-moral” are perhaps morally indifferent, but some of them may well be immoral. And that brings home the illegitimacy of the Subordinationist’s request. What the requester is asking for is a justification that potentially puts morality at the service of immorality. Depending on one’s conception of morality, this may or may not be ultimately legitimate; it’s unclear whether morality entails that evil may not be done that good may come. But if it’s unclear, it’s at least
possible that morality is not the kind of thing that can ever serve immorality. Suppose that “good” and “evil” are incommensurable, and that evil therefore may not be done that good may come. Then given what morality is, there can’t be any legitimate balancing act between good and evil (or rightness and wrongness); there is no conceivable principle that will provide the relevant ratios for that calculation. If so, then the Subordinationist’s question is truth-blocking with respect to morality.

Where the Explanationist ignores the autonomy of morality, the Subordinationist ignores its authority. Authority is a matter of hierarchy, and morality has to stand at the top of any hierarchy of practical concerns. In subverting this hierarchy, the Subordinationist guarantees the failure of her enterprise: nothing that she succeeds in justifying will retain its identity as morality. Notice, again, the defender of morality’s authority needn’t argue at great length for the truth of his claim. He can simply point out that if it is true, the Subordinationist’s question will block access to that truth by the very nature of the question. The potentially truth-blocking nature of the Subordinationist’s question casts doubt on its propriety, unless the Subordinationist can arrive at a conception of morality that is compatible with making it a means to non-moral ends. That, however, seems to run afoul of morality’s impersonality, and seems at least prima facie, a hopeless task.

4.6 The Case for the Dilemma: Premise (3)

The truth of premise (3) seems relatively trivial by comparison with the preceding two. We know what the moral point of view is, so we know what it is to accept it. The moral point of view is defined in such a way as to make clear what it would be to reject
it. That leaves three possible attitudes besides acceptance and rejection: ignorance, neutral skepticism, and indifference. A person might not know enough to be able to accept or reject the moral point of view; a person might be undecided about accepting or rejecting it; or a person might actively withhold acceptance or rejection. The anti-foundationalist will want to argue that none of these three possibilities is coherent. Let’s take these in turn.

With respect to ignorance, it’s hard to see how a person could be ignorant of the nature of the moral point of view, and intelligibly pose the question “Why should I be moral?” in the first place. The question itself presupposes that I have some grasp of the concept of morality. If my conception of the moral point of view is vague or fuzzy, then so will my corresponding sense of the question “Why be moral?” The more vague the sense, the less clear it is that I’ve asked a genuine question in the sense specified above.

Skepticism and indifference are somewhat subtler cases. Since similar considerations apply to them, we can treat them together. On the skeptical supposition, we’re to imagine a questioner $Q$ who knows the difference between the moral and non-moral points of view, but simply can’t decide between them. On the indifference supposition, we’re to imagine some $Q$ who knows the difference between the two points of view but will not decide between them.

Two things might be said about these cases. First, it’s impossible to imagine someone’s being either skeptical or neutral about the two options for very long: skepticism or neutrality seems impossible to live in practice, if only because life itself forces choices on us, and if it does, the choices we make will fall into “moral” and “non-moral” categories by the very definitions of those categories. But second, even if we
conceded that a person could adopt a relatively fleeting epistemic neutrality—or alternatively, if we insist that the impossibility of living skepticism is irrelevant to the issue—it’s still hard to imagine an agent who adopted skepticism of the relevant kind on literally non-moral grounds. It takes effort to maintain and motivate skepticism; most forms of it have been motivated by quasi-moral considerations. To the extent that skepticism derives from some motive, we can categorize that motive as moral or non-moral, and so the skeptic faces a kind of self-reflexive incoherence.

4.7 Conclusion

If this is right, and there are no further cases, then the choice between accepting or rejecting the moral point of view is what William James called a “live option”: living, forced, and momentous. Though we’re (ex hypothesi) free to choose one or the other, we have to choose one or the other, and either way, the choice makes a significant difference to how we subsequently view the world and act. If this is true, then premise (3) is true, and we can take its truth to be equivalent to the affirmation of the antecedents of premises (1) and (2). Having vindicated those premises, and taking the dilemma to be valid, the premises yield the conclusion, and the dilemma is vindicated. If so, the Question is either pointless or illegitimate in the senses described above. Hence the foundationalist project is misconceived, and we are pushed to anti-foundationalism about ethics.
4.8 Notes

1 Throughout this chapter, for ease of exposition I take “foundationalism” to denote meta-ethical foundationalism as described in Chapter 3.


3 It’s far from clear that proponents of this view have a particularly strong thesis to offer even about totalitarianism. Berlin and Rawls make their claim on the basis of sheer unsupported assertion. Arendt’s account, though more obviously based on historical fact, offers very little in the way of argument to tie those facts to foundationalism, and equivocates constantly between theories that offer comprehensive explanations and those that make comprehensive prescriptions.

4 For instance, a view like Peter Unger’s, as expressed in Peter Unger, Ignorance: A Case for Scepticism (Oxford: Oxford University Press, 1975).


6 The standard move is to take egoism as a representative form of foundationalism, and then to take a rebuttal of egoism to rebut foundationalism. The locus classicus, of course, is H. A. Prichard, “Does Moral Philosophy Rest on a Mistake?” Mind, vol. 21, no. 81 (1912), pp. 21-37. For some early post-war examples of


The list could easily be extended to the present. See also references in note 6 above.

Not all questions start with these, of course; some begin with auxiliary verbs, but I think the point made in the text covers question of that sort as well.

It’s not that easy to come up with pure examples of this for two reasons. First, one can always stipulate that the tacit conditions of asserting some question were opaque to some questioner in some context, and there is no way to rule this out entirely.

Relatedly, complications arise with respect to what we might call meta-questions and use/mention issues embedded within questions. Consider the question, “Why ask questions?” This may at first seem pointless, because one is tempted to think, perhaps, that if one couldn’t ask the question without knowing the answer, one wouldn’t ask questions unless one knew why questions were asked. But that’s not necessarily so: it’s not pointless to ask a question, and then to wonder at a meta-level about the reasons for (or perhaps the causes behind—or both) that activity. One might be articulating one’s initially inchoate sense of puzzlement.

Or consider Wesley Salmon’s question, “Why ask ‘why’?” This is actually the very same question as “Why ask questions?” (Wesley C. Salmon, “Why Ask ‘Why’?” in Salmon, *Causality and Explanation*, p. 125); the difference is merely one of grammatical form. But the difference in grammatical form makes the use/mention issue explicit, and thereby makes explicit the point of the question. This is especially the case if the mentioned ‘why’ has a different sense from the sense in use, as it does in Salmon’s example: Salmon’s question is in fact “Why everyday ask ‘why scientific explanation’?” But complications aside, I think the general idea of a pointless question is clear enough.

By “justificatory,” I mean that moral norms justify actions or persons, not that morality as such is in need of a justification.


It should be obvious from the foregoing that the preceding account of conventional morality is biased against certain versions of meta-ethical foundationalism, most notably ethical egoism.


This is the thesis of Walton, *Begging the Question*.

I would ideally have called these questions “reductionist” questions, but having used the term “reductionist” in a different context early in this study, I settle for “explanationist” as a second-best option.


I thank Neven Sesardic for helpful conversations on these issues.

The guiding/goading distinction is from W. D. Falk, “Guiding and Goading,” in *Ought, Reasons, and Morality: Collected Papers of W. D. Falk*, ed. Kurt Baier (Ithaca,


28 Thus Descartes describes the motivation for his method of systematic doubt throughout Meditations 1 as a moralized concern for truth. I thank Daniel Jacobson for raising this issue.


30 Thanks to Carrie-Ann Biondi for helpful editorial feedback on this chapter, and thanks to an audience at The College of New Jersey (April 1999) for useful discussion.
CHAPTER 5:

THE CASE AGAINST THE EROTETIC DILEMMA AND A NEW QUESTION FOR
META-ETHICAL FOUNDATIONALISM

5.1 Introduction

In Chapter 4, I laid out the erotetic dilemma, which takes issue with meta-ethical foundationalism at the most basic level, challenging the question that gives rise to the problem that meta-ethical foundationalism takes itself to solve. By the terms of the erotetic dilemma, if there is no such question, meta-ethical foundationalism cannot get off the ground. The success of any particular version of foundationalism is thus rendered moot: a theory cannot succeed if, as the erotetic dilemma suggests, it has no task at which to succeed.

As I suggested in Chapter 4, the erotetic dilemma is obviously valid, and a century’s worth of philosophers have thought its premises plausible enough to think it sound. In this chapter, however, I want to suggest that its plausibility is an illusion. The plausibility of the erotetic dilemma derives from a set of long-entrenched assumptions about the nature and definition of morality—that is, about what counts as an instance of the concept “morality.” These assumptions, seldom if ever argued for, set the background beliefs which regulate judgments of plausibility about the nature of morality, and in so
doing, constrict the possible interpretations of the “Why be moral?” question. In other words, the dilemma seems plausible only on a highly contestable but seldom contested conception of the concept of “morality”—and by implication of moral agency and justification. Once we identify these assumptions and put them in question, I’ll argue, the dilemma simply falls apart.

Once it does, we can re-formulate the “Why be moral?” question so as to give us a non-trivial question that evades the erotetic dilemma while motivating a foundationalist approach to ethics. In the last section of this chapter, I’ll defend a specific formulation of the question: “What is the justification, compatible with internalism and metaphysical libertarianism, for why a rational agent must rank alternative courses of action and act on the rankings that he or she sets?”

5.2 The Complexity of the Question

The erotetic dilemma, to repeat it from the preceding chapter, runs as follows.

Suppose we have a questioner, \(Q\), asking some version of the Question, “Why should I be moral?” Then

1. If \(Q\) asks the Question from within the moral point of view, the Question is pointless.
2. If \(Q\) asks the question from outside the moral point of view, the Question is illegitimate.
3. \(Q\) can only ask the question from within or outside of the moral point of view.
4. Hence the Question is either pointless or illegitimate.
It’s tempting, on inspecting an argument of this sort, to evaluate it by a mechanical application of tests for validity and soundness. In that case, we would look at the form of the argument, see that it instantiated a valid argument form (constructive dilemma), and then proceed to evaluate each premise, checking to determine its susceptibility to counterexamples.

But there is another, and in my view, more pertinent and fundamental way of evaluating the argument, which is to offer an analysis of the concepts that constitute each premise. A conceptual analysis of this sort could challenge the argument in one or both of two ways. For one thing, it could undermine the argument’s claim to validity by exposing equivocations in terms common to two or more premises: a prima facie valid argument is rendered invalid if the same term means different things in different premises. Second, it could reveal the fact that the dilemma derives its plausibility from specific and contestable conceptions of the concepts expressed in the question. A question can only seem pointless if we assume that it has a transparent answer, and illegitimate if we know a priori that the formulation of the question blocks access to some crucial truth. In the present context, “Why be moral?” can only seem pointless if we assume that it is obvious why we should be moral. It can only seem illegitimate if we know a priori that the act of asking the question is truth-blocking with respect to the nature of morality.

I want to suggest that the erotetic dilemma simultaneously falls prey to both sorts of objection. The dilemma relies for its plausibility, I’ll argue, on highly contestable interpretations of the terms that make up the “Why be moral?” question. In so doing, it fails to take stock of the equivocality of those terms. If we contest the assumptions behind the anti-foundationalist interpretation of the question, we see not only that the
assumptions are undefended and unfounded, but that the question is susceptible of a
gigantic variety of interpretations that far outstrip the resources of the dilemma. If so, the
dilemma applies only to certain construals of the question, leaving other construals
untouched. The untouched construals can thus motivate varieties of meta-ethical
foundationalism immune to the erotetic dilemma.

In a nutshell, the erotetic dilemma rests on three assumptions (or four, depending
on how one counts them), as follows.

1. “Morality” is a transparent concept constituted by “the moral point of view,”
which is itself a transparent concept. It is more or less axiomatic that specifically
moral norms are categorical in binding force and impersonal in character.

2. We can address the question—and generally, do moral philosophy—in the
absence of any deep account of the nature and content of human interests. It is, in
effect, a pre-theoretical datum that “interest” and “self-interest” are concepts with
a transparent meaning. It is likewise a pre-theoretical datum that self-interest is
incompatible with morality because if followed consistently, self-interest leads to
immorality of some kind (predation, narcissism, etc.). Likewise, since
impersonality entails universalizability, and self-interest is incompatible with
universalizability, self-interest is incompatible with morality.

3. When we do the foundations of ethics, what we’re looking for is a justificatory
account as opposed to a causal one. “Justification” and “causal explanation” are
mutually exclusive tasks.

Given the preceding, “Why be moral?” is a relatively simple question, hence easily
demonstrated to be susceptible to the erotetic dilemma. Given the preceding assumptions,
the options—asking the Question within versus outside of—the moral point of view are clear, unequivocal, exhaustive, and exclusive. Since they are, there are no options left.

The assumptions just described give the erotetic dilemma its plausibility because they seem to constrict logical space in a way that makes the alternatives presented by the dilemma seem to be the only possible ones. One obvious strategy for undermining the dilemma, then, is to challenge their implicit claim to having exhausted logical space. To undermine the dilemma, in other words, we have to add complexity to the Question—or recognize the complexity latent in it—which anti-foundationalists ignore or seek to delegitimize. Generally speaking, the greater the number of conceptual possibilities we uncover or open up, the more complex the Question becomes. The more complex the Question, the greater the number of versions of the Question we can generate, and the greater the number of versions, the higher the likelihood that the dilemma fails for neglecting some version of the Question. By the same token, the greater the number of Questions, the higher the likelihood that the dilemma commits a fallacy of equivocation for failing to distinguish relevant versions of the Question.

Some simple arithmetic should help us see the problem facing the anti-foundationalist. On the interpretation of the Question presented in Chapter 4, and accepted by most writers in the literature, we can generate four versions of the Question, as follows:

1. Pointless Question: Why should I be moral?
2. Subordinationist Question (illegitimate): Why should I be moral?
3. Explanationist Question1 (illegitimate): Why should I be moral?
4. Explanationist Question2 (illegitimate): Why should I be moral?
Question (1) falls prey to the first horn of the erotetic dilemma, and Questions (2)-(4) fall prey to the second. But a simple thought-experiment should show how easy it is to generate new versions of the Question.

We generated the preceding four versions of the Question by assuming that “why” has two senses (justificatory and explanatory), “should” has two senses (moral and non-moral) and morality has one sense (morality\(_c\)). The fundamental assumption that drives the dilemma is of course the content of morality\(_c\). But suppose that we discovered or constructed an alternative to morality\(_c\). Putting things very conservatively, that would give us two competing senses of morality, call them morality\(_1\) and morality\(_2\). These two senses of morality would correspondingly generate two new senses of the specifically moral should—call them should\(_{m1}\) and should\(_{m2}\). Suppose further that we could generate a third, distinct sense of “why,” reducible neither to why\(_j\) or why\(_e\); call this why\(_r\) for the “revisionist conception of ‘why’.” And suppose, finally, that we generated different conceptions of moral agency or moral psychology of relevance to the content of morality, and therefore of relevance to the “I” asking the Question; call these I\(^*\) and I\(^**\). That would, in addition to Questions (1)-(4) above, give us (at least) eight more versions of the Question:

5. Why\(_r\) should\(_{m1}\) I\(^*\) be moral\(_1\)?
6. Why\(_r\) should\(_{m2}\) I\(^*\) be moral\(_2\)?
7. Why\(_r\) should\(_{m1}\) I\(^**\) be moral\(_1\)?
8. Why\(_r\) should\(_{m2}\) I\(^**\) be moral\(_2\)?
9. Why should \( m_1 \) I* be moral_1?

10. Why should \( m_2 \) I* be moral_2?

11. Why should \( m_1 \) I* be moral_1?

12. Why should \( m_2 \) I** be moral_2?

Granted, this is so far a purely hypothetical claim involving merely nominal possibilities; we would want to see whether we could generate real possibilities here, and we would want to investigate the plausibility of those possibilities. In working through them, we might well find that some were senseless, others implausible or redundant, or even that all of them somehow fell prey to the erotetic dilemma. That, however, would require working through them; it’s not something we can assume from the outset. In any case, the relevant point is entirely general: arithmetic increases in senses of the constituent terms of the Question greatly multiply the conceptual complexity of the Question. If we merely add one more sense of “why” to what we currently have, we get eight more questions; likewise if we add one more sense of “I” or of morality. If we simultaneously add one more sense of “why,” of “I,” and of morality, we add twenty-four versions of the question to our original list. It’s not hard to see how easy it is to add mind-boggling complexity to the Question.

Assuming that we can convert these nominal possibilities into real ones, we create two problems for the anti-foundationalist, one minor and one major. The minor problem is quantitative and so to speak, bibliographical. If we have twenty-four versions of the “Why be moral?” question, the anti-foundationalist needs to run the erotetic dilemma on far more versions of the Question than might have been anticipated. That’s enough busy
work for a decade, and bibliographically speaking, it is busy work that has been left undone.

The bigger problem is qualitative and conceptual. As we generate questions on this schema, we don’t merely increase the sheer number of questions, but create a sort of multi-dimensional conceptual novelty. The problem isn’t merely that the anti-foundationalist has to churn through more questions than he otherwise had to; it’s that he now has to confront entirely new versions of the problem than he previously confronted. If there really is, let’s say, a conception of “why” that is reducible neither to why_j or why_e, it’s obvious that the anti-foundationalist will need a new sort of argument to undermine versions of the question based on that new conception of why: if why_r is reducible neither to why_j nor why_e, then arguments against “Why_r should I be moral?” will have to be based on considerations that are themselves not reducible to the considerations against Why_j or Why_e. Likewise with alternative senses of “moral” or of moral agency. In generating versions of the Question, we’re not merely generating generically similar versions of the same Question, but wildly different sorts of Question that introduce philosophical issues of a much broader sort. Assuming we can succeed at that task, we should expect subtle changes in the meaning of the Question, and by implication, changes in the very nature of the relevant topics.

We can, I think, convert the nominal possibilities into real ones by contesting three anti-foundationalist assumptions—the first about morality (and “should”), the second about moral agency, and the third about the relation between explanation and justification.
First, “morality” is a highly contested concept. Anti-foundationalists dramatically underestimate its complexity, and so operate with a highly impoverished conception of its content. Contrary to a wide consensus, it is not at all obvious what counts as “a morality,” or of taking “the moral point of view.” If so, the complexity of the Question is going to be a function of the complexity of the concept of “morality,” and will mean different things relative to different conceptions. The greater the complexity, the more it means; the more it means, the less likely that the Question is pointless.

Second, much of the support for the authority of morality comes from the assumption that morality is (to some degree) incompatible with self-interest. Morality is typically conceived in the literature as though it were simply a device to counteract self-interest or to resolve conflicts of interest. But the philosophers who make this assumption operate with a conception of self-interest for which they offer no support whatsoever. If we contest it, egoism becomes a candidate morality theory, and we’re obliged to re-conceptualize the usual way of understanding morality—and by implication the “subordination” of morality to non-morality. The more radical the revision, the less likely that the Question is illegitimate.

Third, contrary to anti-foundationalist assumptions, justification and explanation are related to one another in highly complex ways that undermine the notion that they are jointly exhaustive and mutually exclusive categories. The more decisively they undermine that notion, the less likely that the Question is either pointless or illegitimate.
5.3 The Moral Point of View: Categorical versus Conditional Conceptions of Moral Obligation

5.3.1 Categorical and Conditional Conceptions of Moral Obligation

As remarked in Chapter 4, one of the defining commitments of morality, and by implication the anti-foundationalist conception of the moral point of view is the claim that moral norms are categorical in their binding force. Though the idea derives from Kant’s conception of a categorical imperative, it is characteristically detached from that view, and for present purposes, I follow what I take to be Rawls’s authoritative treatment of the notion.¹ A categorical moral requirement is a requirement that “applies to us”—binds us—as persons, independently of any voluntary act we perform, performative or otherwise. As a requirement, it is something we must do such that failure to perform it by definition licenses a judgment of immorality and justifies appropriate action against us. As a requirement that binds us as persons, its binding force bears no necessary connection to institutions or social practices; categorical obligations “hold between persons irrespective of their institutional relationships; they obtain between all as equal moral persons.”² But their categoricity consists, specifically, in binding us independently of what we do. We do nothing to incur them; we simply have them apart from anything we do or don’t do. Thus we have them regardless of our desires and choices, regardless of how we act, and regardless of whether we choose to act or not. We might say that they bind us in virtue of our sheer existence.

As just defined, a categorical moral norm is (to use Rawls’s terminology) one that is unconditional—conditional neither on the structure of society nor on the individual performance of acts. Logically, then, we could conceive at least three alternatives to a
categorical moral norm corresponding to three types of conditionality: a norm could be
conditional on the structure of society, conditional on an individual’s voluntary actions,
or both. Philosophers have defended all three sorts of view, but (following Rawls again) I
think it makes sense to think of the essential feature of a categorical moral norm as its
independence from individual voluntary acts. What makes something categorical is that it
is not conditional on the person it binds.

If that’s right, we may define a \textit{conditional moral norm} as a moral requirement
that applies to us as persons, but does so precisely in virtue of a voluntary act or set of
actions, performative or otherwise, whereby we incur the obligation. Understood in this
way, a conditional moral norm is conditional on one and only one thing: a certain kind of
act. That is to say: on the conditionalist conception, \( S \) incurs moral obligations if and only
if \( S \phi' \)'s, where ‘\( \phi' \)’ denotes some action type, and performance of a token of that type is
such as to put oneself under a moral obligation. Notice that being independent of social
structures, a conditional moral obligation will bind anyone who performs \( \phi \), regardless of
anything else about him. By implication, of course, such an obligation will not bind
someone who does not \( \phi \). In principle, then, on the conditionalist conception, \( S \) can escape
the clutches of morality so long as he can manage to avoid \( \phi \)-ing.

As a paradigm example of a conditional moral norm, consider promising (and
owing): I incur a debt to deliver on a promise if and only if I voluntarily promise to do
something. Correlatively, I can escape being governed by promise-norms if I assiduously
avoid making any promises. On a conditionalist conception of morality, however, what is
ture of promises is true of morality across the board: I incur moral obligations if and only
if I voluntarily do something. The basic task of a conditionalist moral theory is to identify that something in a defensible way.

By and large, contemporary philosophers have taken it for granted that morality must be categorical. Sporadic exceptions aside, few have taken the conditionalist conception of morality very seriously.³ And so the categorical conception has by and large been taken for granted as constituting the moral point of view, and defining morality. Darwall’s view is representative of the mainstream consensus:

Many of us share the Kantian conception of at least a part of morality as embodying requirements and recommendations for action which, at their most basic and general level, are common to any being who is capable of recognizing and regulating his life by them. Many, (though assuredly not all), see these considerations as reasons for us to act; reasons whose existence depends simply on the validity of the moral considerations themselves. The deep attraction of this view is shown by the fact that a perennial and seriously defended position in moral philosophy is that it is impossible even to raise the question whether there is reason to be moral.⁴

Notice Darwall’s claim that “at least a part of morality”—the phrasing seems to imply “the most important part”—embodies requirements “common to any being…capable of recognizing and regulating his life by them” (my emphasis). The italicized word conveys the thought the morality binds us in virtue of our capacities for action, not the actions themselves—the essential feature of the categorical conception of morality.

The bottom line, then, is this: The erotetic dilemma turns on the claim that morality exhausts logical space—that morality just is morality, and there are no alternatives to it. The categoricity of moral norms is essential to morality. Hence if we can successfully put the arguments for categoricity into question, we can put into question the anti-foundationalist’s claim that morality exhausts the landscape of possible
conceptions of morality. As far as I can see, the arguments for categoricity (insofar as they count as arguments, which is unclear) fall into two categories, neither successful. I take them in turn.

5.3.2 The Linguistic Argument for Categoricity

A first argument is less an argument per se than a blank assertion. The assertion consists of the quasi-linguistic or sociological claim that “our” ordinary language reveals a commitment to the categoricity of moral norms. Thus, it is argued, when “we” say, “It’s wrong to murder,” “we” do not mean the wrongness of the act to be contingent on the would-be murderer’s act; “we” mean the wrongness to be unconditional. Likewise, when “we” recognize a duty of justice or benevolence, “we” do not mean duties of this sort to depend on the performance of any sort of act by the person on whom the duty falls. “We” mean that the duty falls on persons regardless of their acts. Since “our” ordinary language reveals a commitment to a categorical conception of morality, and (to fill out a tacit step of this rarely articulated “argument”) ordinary language reliably tracks the truth about morality, it follows that moral norms are categorical in binding force.

Spelled out explicitly, this argument is almost impossible to take seriously. For one thing, the preceding rendition of it involves an ample dose of charity: the argument is rarely if ever spelled out as explicitly as I just have. Second, it’s neither clear as a sociological matter whether it’s true, nor clear how one would go about determining its truth. Third, it’s perennially unclear what the referent of “we” is supposed to be. Indeed, it’s not even clear what language is being referred to. But fourth, even if we suppose the sociological claims to be true, they obviously beg the question. “Our” talking a certain
way doesn’t mean that that way of talking tracks the truth unless we assume, as a general principle, that talk—or talk of a certain kind—tracks the truth. And, to put the point bluntly, no such argument has ever been offered; nor is it even clear what such an argument could conceivably say. At best, what these linguistic claims tell us is that certain entrenched ways of speaking about morality seem to presuppose a categorical conception of morality. But nothing about the argument tells us whether the truth about morality is determining the way “we” speak (if indeed “we” really speak that way), or rather that the way “we” speak is being determined by what “we” take to be the truth about morality. I conclude, then, that this argument is simply hopeless, as are all arguments of its type.  

5.3.3 The Escapability Argument against Conditionality

A second set of arguments represents a development on the linguistic ones just described. If we look carefully at claims about moral obligation, on reflection what should strike us as odd about a conditional conception of moral norms is its failure to capture the inescapability of moral norms. On a categorical conception, moral norms are inescapable: they bind us regardless of how we try to escape them; there is no way to beg off from them. On a conditional conception, by contrast, moral norms are escapable. Since obligation is conditional on S’s φ-ing, S can escape morality by refusing to φ. And that, in turn, seems a problematic or counter-intuitive conclusion. It allows people to shuck their moral obligations with impunity. It gives us no justification for punishing them or holding them accountable. And it holds morality hostage to their mere whims or tastes: moral obligation waxes and wanes with our actions. But surely moral obligation is
a robust phenomenon: it overrides what we want. Indeed, that’s part of what it’s there for. A categorical conception of morality captures this; a conditional one does not.

This argument gets one thing right: categorical moral norms are indeed inescapable. But, of course, that is the problem with them: the inescapability of moral norms is simply asserted as though it were an axiom. But the inescapability of moral norms so conceived leads us directly to their inexplicability and non-justifiability. Categorical moral norms do not get their binding force from the relation they bear to our actions. But just for that reason, it is unclear whence their binding force derives—precisely the question the anti-foundationalist refuses to answer and seeks to delegitimize. This inexplicability seems (or might seem) a tolerable price to pay if we accept the picture that the proponent of categorical moral obligation paints: either moral obligations are categorical, or they are escapable, hence escapable with impunity, subversive of responsibility, and so feeble as to be hostage to our whims and tastes. But every one of these claims is a non-sequitur, turning on the failure to get clear on the distinction between an escapable and inescapable norm.

We should remind ourselves, in semi-formal terms, of the definitions of categorical and conditional moral obligation. Take a norm $N$ and a moral agent $S$, where $S$ is assumed to be capable of following $N$. If $N$ is categorical, $N$ is inescapable for $S$ in the sense that nothing $S$ does will relieve $S$ of the obligation to comply with $N$. By contrast, if $N$ is conditional, $N$ obliges $S$ if and only if $S$’s, for some (as-yet) unspecified value for $\phi$. Suppose \textit{ex hypothesi} that $S$ does indeed $\phi$. Then $S$ is obliged to follow $N$. Suppose that $S$ refuses to $\phi$. Then $S$ is not obliged to follow $N$. On this view, something—
precisely one thing, and only that thing—relieves S of the obligation to comply with N: namely \(~\phi\)-ing.

The contrast between categorical and conditional obligation is that on a categorical view, *nothing* you do can relieve you of the requirement to observe a norm, whereas on a conditional view there is in principle *one* thing you do that can relieve you of that requirement. Note, however, that put in formal terms, the “something you do” that relieves you of the requirement on the conditional account has been left entirely unspecified. It’s crucial to remember that *its* identity is what determines the meaning of “escapability” in a conditionalist conception of morality. What defenders of a categorical conception have failed to see is that there are values for \(\phi\) which, though strictly speaking escapable, are not at all easy to escape. Indeed, there are values of \(\phi\) which are escapable at so colossal a price as to render them close to inescapable (without literally being so).

A somewhat fanciful example should drive this home. Imagine a conditional conception of morality whose norms were contingent on one action: breathing. On this view, if S breathes, S falls under the jurisdiction of moral norms, but if she doesn’t, she is not so governed. Strictly speaking, this view makes morality perfectly escapable. If S wants to escape the reach of morality, “all” she needs to do is to stop breathing. If she does indeed stop breathing, she is (*ex hypothesi*) beyond good and evil—in fact, beyond all moral judgments. We might well say that the choice to breathe or not to breathe is itself beyond moral judgment on this view. But escapable as morality is on this view, it seems absurd to dwell on its escapability, or treat its escapability as any great liability. One way to escape morality on this view might be to stop breathing entirely. But of course, the price of escape would then be loss of the agent’s existence and identity.
Perhaps some persons might elect to escape morality at this price. We might then regard it as a liability of this view that it cannot judge their electing to escape morality. But it’s not clear it is a liability. We might simply reply that morality is for the breathing.

One might object that the “morality-conditional-on-breathing” view still has a baleful consequence. Suppose $S$ wants to commit some immorality, where the immorality would take just a few seconds to effectuate. Suppose that $S$ holds his breath during those few seconds, performs the action, and then begins once again to breathe. Hasn’t $S$ escaped morality?

No. There is, for one thing, a question about how to individuate the relevant action. It’s implausible to think that actions can be so finely individuated that there can be an action of a few seconds’ duration initiated when a person was not breathing that was not in some way continuous with a predecessor action initiated when the person was breathing. So long as the duration of action-while-not-breathing was causally dependent on prior action-while-breathing, we could say that the identity of the action was determined by action-while-breathing, in which case the person would (ex hypothesi) be subject to moral obligation. In any case, assuming that the agent began to breathe after the action, he would be subject to moral judgment for the consequences of the action initiated while he wasn’t breathing. In any case, bear in mind that what the conditionalist conception says is that $S$ incurs moral obligations by $\phi$-ing. It doesn’t say—nor does it follow—that moral obligations incurred at $t_1$ by $\phi$-ing can be extinguished at $t_2$ by $\neg\phi$-ing. What follows is merely that moral obligations are incurred by $\phi$-ing at some $t$, and would not have been incurred had $S$ not $\phi$-ed.
Obviously, I’m not defending the thesis that morality really is conditional on the act of breathing. The breathing example should, however, serve to clarify what the proponent of the conditionalist conception is really saying. A plausible conditionalist morality will make moral obligation conditional on an action which, qua action, is by nature an obligation-incurring act, but which is also by nature so ubiquitously performed that “escapability” will (essentially) be a moot point.

If this is right, then the arguments advanced by the proponent of the categorical conception all miss the point. It will be true on a conditionalist view that we can escape the authority of morality by refusing to \( \phi \). But if \( \phi \)-ing is like breathing, the price of escape is so high as to obviate the threat of it. Nor can persons really escape morality with impunity: asphyxiation and death are not impunity. Nor does it remove our grounds for dealing accordingly with those who try to escape from morality: they might well try to escape, but morality might well give the rest of us advance directives for dealing with those who make the attempt. Nor does it hold morality hostage to “whims and tastes”: breathing is neither a whim nor a taste. Nor can one legitimately say that morality’s binding force is rendered non-overriding. So long as one is breathing, morality might well override all other considerations. The only threat to morality’s authority, insofar as there is one, comes from those who adamantly refuse to breathe. But this is just to say that morality is conditional on one of our natural needs. Those who lack such needs are beyond morality. But they are also irrelevant to human life.

Consider another, less fanciful example than the preceding one loosely modeled on Gewirth’s derivation of the Principle of Generic Consistency in *Reason and Morality*.7 Imagine a conditionalist interpretation of morality according to which \( S \) incurs moral
obligations if and only if he performs some (but any) intentional action as such. When S acts, S is governed by morality, but S can escape morality by refusing to act. A proponent of such a view would, to be sure, have to “concede” (if that’s the right word) that moral obligation was on such a view escapable. But if on this view, action is the catalyst for obligation, and inaction the price of escape from obligation, it seems obvious that the escapability argument against the conditional conception of moral obligation proves much less than its champions have thought it did.

5.3.4 Conditionalist Conceptions of Morality as a Rival to Morality_c

Let’s suppose, as I’ve suggested in the preceding two sections, that the arguments for thinking of moral obligation as inherently categorical fall short of the mark. We could in principle infer that moral obligation is in fact conditional in binding force, but I propose to make the weaker inference that the conditionalist conception of morality is at least a worthy rival of the categorical conception, and so an alternative to morality_c. If so, morality_c is not, as has so often been assumed, the only game in town. There are at least two—morality_c and what I’ll call a conditionalist conception of morality. This contrast serves to weaken the force of the erotetic dilemma. But it also throws some light on the meaning of the “Why be Moral?” question that would otherwise seem obscure.

As Darwall suggests in the passage quoted in section 5.3.1 above, if we adopt morality_c, we ipso facto endorse the categorical conception of moral obligation, and in so doing render the “Why be moral?” question unintelligible. If categoricity is (by fiat) definitory of moral obligation, then to know what moral obligation is, is inherently to know that it is categorical. But if categoricity is irreducible and inexplicable, it makes no
sense to ask for a further justification for it, which is precisely what the “Why be moral?” question does. In that light, the erotetic dilemma may well make sense: the Question will either seem pointlessly naïve or problematically illegitimate.

If, by contrast, a conditional conception of morality is credited as an option, the Question immediately becomes more intelligible as an inquiry into the binding force of morality. If morality could be either categorical or conditional, two questions are which one it is, and how we know. If it turns out to be categorical, the justification why (insofar as there is one) should be contrastive: it should explain in non-trivial or tautological terms why it cannot be conditional on anything. If it turns out to be conditional, we need to know why it is conditional (as opposed to categorical); why it is conditional on ing (as opposed to something else); and how it handles questions of escapability. In neither case do those questions seem either pointless or illegitimate, and if not, neither is the Question.

5.4 The Moral Point of View: Egoistic versus Impersonalist Moralities

5.4.1 Arguments against Egoism’s Claim to Being a Genuine Morality

A second defining commitment of morality is the incompatibility of specifically moral norms with norms of prudence or self-interest, and by implication the incompatibility of the specifically moral point of view with that of prudence or self-interest. A weak version of this view is agnostic about the extent to which self-interest conflicts with morality, and contents itself with the claim that moral norms, qua moral, are incompatible with egoistic ones qua egoistic. A stronger version asserts the weak view but goes on to allege a fundamental conflict between morality and self-interest. On
the stronger views, the very function of morality is to counteract self-interest. The idea here is that untrammeled self-interest ubiquitously practiced would lead to gross immorality and/or disorder; while self-interest cannot literally be abolished, morality functions as a side-constraint on it so as to keep its effects firmly in bounds.

The general thesis is that egoism is incompatible with morality, so that specifically egoistic norms are incompatible with the moral point of view. The list of particulars against egoism can, as I see it, be reduced to four criticisms, each of them intended to distinguish the specifically moral from the specifically egoistic, and each of them a variation on the same theme.\(^8\) I’ll list the arguments in serial order, offer a general response to all of them at once, and then enumerate specific responses to them in turn.

The first argument is that unchecked egoism leads ineluctably to immorality via predation or malevolence. The argument here is that certain uncontroversially beneficial goods are best secured, at least on occasion, by predation. If so, egoism has nothing adverse to say about predation so motivated. Likewise, some agents may have radically malevolent conceptions of their interests. Since egoism tells them to act on their interests, egoism has no way of criticizing malevolence.\(^9\)

The second argument is related to the first: in leading to immorality, unchecked egoism would lead to social disorder. Egoism tells everyone to act in their self-interest. But if self-interest leads (as per the preceding argument) to predation and/or malevolence, egoism leads not only to immorality but to a state of affairs that resembles Hobbes’s description in *Leviathan* of the “state of nature.” Since that’s obviously suboptimal, egoism fails at what we might regard as a basic task of a moral code, namely, giving us
action-guiding advice that allows us to coordinate our actions and avoid or resolve fundamental conflict.¹⁰

A third argument is that egoism is insufficiently other-regarding. Even if egoism is less predatory or malevolent than the preceding two objections suggest—even if it isn’t predatory or malevolent at all—the fact remains that egoistic self-regard is incompatible with the genuine other-regard that morality requires of us. An egoist, at bottom, is out exclusively for himself. But love, friendship, benevolence, sympathy, empathy, and care require genuine other-regard. Since egoism is incompatible with genuine other-regard, it is incompatible with love, friendship, and the like. But a putatively moral conception incompatible with these things is not one worthy of the name. Hence egoism cannot count as a genuinely moral conception.¹¹

A fourth argument is a more general and rationalist version of the third.

Impartiality and objectivity would seem to be minimal necessary conditions for a moral conception. Impartiality, in turn, entails universalizability: if a norm cannot be universalized, it can scarcely count as a norm, i.e., a general principle, and so fails the minimal test for being a moral norm. Objectivity requires that we not be judges in our case, for when we are, we are incapable of distinguishing rational judgment from the heat of passion, and so cannot be said properly speaking to be capable of being governed by any norms. But egoism (as Moore was the first to argue)¹² is incompatible with universalizability, and so with impartiality. Similarly egoism (as William James argued)¹³ requires us always to be judge in our own case, and so violates objectivity. Again, egoism fails as a genuinely moral conception.
5.4.2 Responding to Arguments against Egoism’s Status as a Genuine Morality

The general response to all four of these objections is that they rest on specific assumptions about the content of self-interest. But in canvassing the broad sweep of literally a century’s worth of such criticisms—from Moore to the present—I found not a single case in which the assumptions about the content of self-interest were defended in any sustained way, and precious few cases in which the assumptions were even made explicit. Philosophers seem to have argued against self-interest as though it were obvious to everyone what counts as an “interest,” and so, obvious how morality and self-interest were incompatible. But it’s not obvious at all. If defenders of morality bear the burden of proof for asserting the incompatibility of morality with self-interest, it seems fairly obvious that they bear the burden of proof in demonstrating this. To demonstrate the incompatibility of X with Y presupposes that we know that X and Y are distinct and, more fundamentally, what they are. It can in my view confidently be asserted that no academic philosopher in the English-speaking world during the twentieth century has even attempted such a thing. It follows that no such philosopher has established—or properly speaking even tried to establish—the vaunted incompatibility of morality and self-interest. Once we see this, the responses to the preceding objections all become variations on a rather obvious theme.

The first argument tells us that egoism leads to immorality via predation or malevolence. The claim that predation and/or malevolence are in our interests is assumed without argument. Being assumed without argument, we can dismiss it without argument. Let us assume instead—merely ex hypothesi—that neither predation nor malevolence are in fact in our interest. Assume, rather, that a principled commitment to justice and
benevolence are in our interest. In that case, the objection doesn’t get off the ground. For if justice is constitutive of our self-interest, then it is simply false that some goods are “best secured by predation.” In fact, the reverse will be true: no good is best secured by predation; goods are best secured through justice. Again, while some agents may have radically malevolent conceptions of their interests, if in fact benevolence is in our interests, they will simply be wrong—as wrong as someone who insisted that cigarette smoking or junk food were healthy or nutritious.

The response to the first argument gives us a response to the second. If egoism counsels justice and benevolence, then unchecked egoism would lead to a just and benevolent form or order, not disorder. And so the second objection fails.

The third argument tells us that egoism is insufficiently other-regarding. If we are “out for ourselves,” it says, we cannot be genuinely concerned with our intimates, friends, or the other objects of our benevolence. This view gets whatever psychological plausibility it has (for those who find it plausible) by picturing the egoist as a sort of narcissist, too wrapped up in his own concerns to extrospect on the world outside of him or pay attention to others besides himself. Let us suppose, *ex hypothesi*, that narcissism of the just-described sort is a psychological disorder, and that psychological health, like physical health, is one of our most fundamental interests. In that case, egoism counsels that we do what it takes to achieve health, and do what it takes to avoid psychological disorder. Thus an egoist in her healthy state would strive not to be (and succeed at not being) a narcissist.\(^{15}\)

Suppose further that our lives go better if we have intimates and friends, and if (as per the preceding two paragraphs) show benevolence in appropriate ways to appropriate
persons under the appropriate circumstances. In that case, friendship, love, benevolence, and the like would all be in our interest. If health is in our interest, and friendship (etc.) are in our interest, and the latter require genuine other-concern, the only obstacle to egoism would be the possibility that it is logically or psychologically impossible simultaneously to be self-regarding and other-regarding, and other-regarding in virtue of one’s self-regard. It is completely unclear (to me) what this incompatibility is supposed to be, and very little in the literature succeeds in clarifying it to any serviceable degree. It’s worth remembering that Aristotle (among others) suggested that self-regard was the best possible basis for other-regard. So sheer appeal to intuition will establish nothing here; Aristotelian intuitions will incline toward some form of egoism.

The fourth objection involves two claims, first that egoism is incompatible with universalizability, and secondly that it’s incompatible with objectivity. It seems to me that neither claim is defensible. Let’s take these in turn.

Universalizability is the principle that if $S$ asserts that an action or policy $X$ is right in circumstance $C$, then $S$ is committed to thinking that $X$ is right in every circumstance relevantly similar to $C$. Applied to egoism, this yields: if an egoist asserts that egoism is right for him, then he is committed to thinking that egoism is right for everyone relevantly similar to him. Let’s take “relevantly similar to him” to mean all moral agents. Then: if an egoist asserts that egoism is right for him, then he is committed to thinking that egoism is right for all moral agents. Fair enough; but why, exactly, is this supposed to be a problem for egoism?

Moore’s objection holds that the egoist, in taking his own interest to be ultimate, is enjoining himself to pursue his own interest. In so doing, Moore suggests, egoist $E$
would have to be enjoining everyone else to pursue his own interest as ultimate, or making everyone’s interest ultimate. But Moore’s objection is simply confused. What egoism asserts as a general principle is that everyone ought to regard the pursuit of E’s interests as ultimate. Being a general principle, it has applications in every individual’s case while remaining a general principle. So egoism doesn’t say that everyone should follow some one person’s interest; it tells everyone to follow their own interest. But that doesn’t lead to the supposed contradiction that Moore asserts, since egoism can without inconsistency regard interests as qualitatively identical but agent-relatively valuable. So while my interest is ultimate in an action-guiding sense for me, and yours is ultimate in that sense for you, we can without contradiction (or even the slightest tension) predicate “ultimacy” of both of them just as surely as we can predicate “health” of two healthy persons while admitting that each one’s health is his or her own.

So Moore’s objection is a non-starter, but it leads to a somewhat more plausible objection (or set of them) regarding recommendations. Suppose that you and I are egoists. Suppose that φ-ing is in your interest, and ~φ-ing is in mine. Suppose, in fact, that it’s true across the board that our interests systematically conflict. How can universalizability apply here? We’re both moral agents. Let us suppose that we are ex hypothesi in relevantly similar circumstances. Egoism tells you to φ, but it tells me to ~φ. Doesn’t that violate universalizability?

This common objection takes for granted what many of the preceding anti-egoist objections take for granted: that it’s intuitively obvious that two moral agents typically are in “relevantly similar” circumstances but have conflicting interests. In the absence of an account of self-interest, there is no reason whatsoever to accept this assumption, and
so no reason to accept the inference about the failure of egoistic universalizability. At best, what this argument tells us is that if egoistic interests were fundamentally conflictual in the preceding sense—if relevantly similar agents had systematically conflicting interests—egoism \textit{would} fail a test of universalizability. But by the same token, the very same argument should suffice to show that if egoistic interests are \textit{not} conflictual in that sense, there is no reason to think that egoism fails universalizability. In any case, the burden of proof surely rests with someone who asserts that similar agents \textit{have} systematically conflicting interests, for if such interests systematically conflict, anyone asserting this should have in hand at least a rough account of what accounts for this systematic conflict. Of course, if the conflict is sporadic rather than continual, we need an account of why the sporadic conflict arises. But the literature contains little to nothing of this nature.

\textit{Mutatis mutandis}, the preceding argument applies to objections to egoism that assert that egoism is incompatible with the practice of giving recommendations. Why, it might be wondered, is it in my interest to recommend that you follow your interest? The objection only gets off the ground if we assume that the interests of egoists acting qua egoists are fundamentally conflictual. If they were, of course, egoism would be incompatible with recommendation-giving. But if they are not, there is no reason to think that egoism is incompatible with recommendation-giving: if I think that interests are harmonious, it can be very much in my interest to enjoin others to pursue their interests, for that way, I benefit through the common advantage.
5.4.3 Egoism as a Genuine Morality: Implications

Though I have done nothing in the foregoing to demonstrate the truth of egoism, the preceding arguments do, I think, suggest (contrary to defenders of the authority of morality), that egoism is a genuine morality. And so, once again, morality turns out not to be the only game in town: moralities can be impersonal or egoistic. This serves, once again, to weaken the erotetic dilemma. The erotetic dilemma assumes for its soundness that morality exhausts the possible range of genuine moralities. But since egoism is a counter-example, that can’t be right. Indeed, if we combine the insights of section 5.3 (on conceptions of obligation) with those of the current section, we get a total of four different competing conceptions of morality:

1. morality: a non-egoistic morality with a categorical conception of obligation
2. non-egoistic moralities with a conditional conception of obligation
3. egoistic moralities with a categorical conception of obligation
4. egoistic moralities with a conditional conception of obligation

If we look more closely, however, we see that the admission of egoism onto the roster of genuine moralities has a series of further (and in a sense deeper) consequences for how we understand the “Why be moral?” question. In fact, it multiplies the iterations of the Question manifold.

5.4.3.1 Optimistic versus Pessimistic Conceptions of the Self

As we’ve seen, one of the issues that divides morality from egoism is that of background beliefs about the content of “self-interest.” These background beliefs, in turn, derive from beliefs about the self whose interests are in question—the “I” asking the
“Why be moral?” question. Following Robert Nozick, we can distinguish what may be called *optimistic* and *pessimistic* conceptions of the self and its interests.\(^\text{19}\)

As Nozick puts it, the optimistic conception holds that there are no objective conflicts of interests between persons. The general idea here, which has antecedents in Aristotle and Spinoza, goes as follows. Moral agents have certain fixed needs, the satisfaction of which, through virtue, constitutes their objective interests. Such interests, rationally pursued, make possible an ideal of intrapersonal harmony over time: moral agents are constituted in such a way as to be capable of living a regretless, non-tragic life, that is, a life in which one’s powers and desires can harmoniously be ordered without its being the case that one faces insurmountable conflicts of principle or desire that significantly mar one’s life. Such conflicts are certainly possible in life, even on the optimistic conception, but occur by either or both of two causes: violations of moral virtue and/or freak and disastrous aberrations in the course of nature. What the conception rules out is the possibility of lifelong intrapersonal conflicts of interest or desire not explained in one of the preceding two ways.

This intrapersonal thesis is related to an interpersonal one: given the relative uniformity of human nature, any two agents in intrapersonal harmony will be in interpersonal harmony. Principled action makes it possible for me to achieve an internal concord of interests; action on the same principles make it possible for you to achieve such an internal concord. If so, the two of us are acting on the same principles, at least at some generic level of similarity, and have the same interests, at least at some level of generality. If so, conflicts of interest between us will be impossible; such interpersonal conflicts as do occur will be traceable to intrapersonal conflicts, which are themselves
traceable to the violations of virtue or “freak occurrences” described above. What this rules out at the interpersonal level is the possibility of irreducible conflicts of interest, i.e., conflicts of interest not explainable by the preceding two principles.

Though Nozick doesn’t discuss a “pessimistic tradition,” we can think of some such tradition as the mirror-image of the optimistic one. Where the optimistic tradition denies the possibility of objective conflicts of interest, the pessimistic one affirms it, insisting not only on the possibility of such conflicts, but on their ubiquity. These conflicts derive from two sources, one of which we can describe as pessimism about depth psychology, and the other as resource pessimism.

*Depth psychological pessimists* hold, contrary to the optimistic tradition, that deeply rooted psychological factors make intrapersonal harmony impossible. On this view, espoused in different ways by St. Augustine, Calvin, Kant, Nietzsche, Freud, and Lacan, moral agents are not constituted in a way that makes internal harmony possible: our capacities and desires are fundamentally disordered, or at war with each other; a commitment to moral virtue can mute the conflicts but never resolve them. Since the optimistic tradition takes intrapersonal harmony to be the basis of interpersonal harmony, the depth psychological objection cuts at the core of the optimistic claim.

*Resource pessimists*, while agnostic about the possibility of intrapersonal harmony, hold that external factors in the physical environment make interpersonal harmony impossible. According to the resource pessimist, even if we could achieve intrapersonal harmony, the optimist assumes too quickly that we can translate our inner psychological achievement into social reality. In fact, the “freak disasters” to which the optimist alludes (and which he cavalierly dismisses as aberrations) are the *de facto* norm
of human life: physical resources are always in scarce supply, and our attempts to exploit them often come at a catastrophically high environmental or social cost. Classical exponents of this view include Hobbes and Malthus; prominent contemporary advocates include E. F. Schumacher, Paul Ehrlich, and others.\textsuperscript{20}

We needn’t decide between these various conceptions to see that they have fundamentally different conceptions of the self and its interests, and so presuppose fundamentally different things when they ask the “Why be moral?” question. The “I” of the optimistic conception is just a different sort of agent from the “I” of the pessimistic conception, and we need to acknowledge this in assessing both the cogency of the Question and the possible answers to it.

5.4.3.2 The Scope of Morality

Recall that one premise of the erotetic dilemma regards the “Why be moral?” question as illegitimate because it seems to subordinate moral considerations to non-moral considerations. If, following the strictures of morality, we think of egoism as a non-moral conception, then egoistic versions of the “Why be moral?” question—questions that ask, in effect, how morality promotes the agent’s self-interest—will seem to have this deficiency. Obviously, if egoism is a moral conception, then egoistic conceptions of the Question will not have this difficulty. Notice, however, that it is essential to the erotetic dilemma that we be able to differentiate moral from non-moral considerations. One interesting implication of an egoistic conception of morality is that this distinction becomes impossible to make—as does the objection founded on it.
My official account of morality, makes no mention of the scope of morality; it depends exclusively on claims about categoricity and impersonalism. But there is, I think, a semi-orthodox view about its scope as well. This view holds that it is axiomatic or obvious that morality governs interpersonal relations (or even more narrowly, a subset of interpersonal relations), leaving the rest of our lives entirely ungoverned by moral principles. The idea, half-borrowed from Kant, is that morality gives us a set of perfect and imperfect duties regarding others. Perfect duties are sharply defined and unexceptional: I have a perfect duty to respect your rights, and little or no discretion about how that is to be done. Imperfect duties are, though equally binding, subject to individual discretion: I have an imperfect duty of assistance to others, but discretion about whom I will assist, at what time, under what circumstances, at what cost. Once we get through the roster of duties to others, on this view, we reach the boundaries of morality. Morality has nothing to say about how we live our lives once our duties are discharged. That is a matter of taste, social convention, or aesthetic, considered in entirely non-moral terms.

This conception of morality sits uneasily with religious conceptions of ethics, which purported to offer comprehensive ways of life to their adherents. This is clearly true of adherents of the Abrahamic faiths; for orthodox Jews, Christians, and Muslims alike, religious ideals and prescriptions bind the whole of one’s life, not merely the interpersonal domain. Indeed, for many religious believers in these traditions, the interpersonal sphere has less ethical significance than an agent’s relation to God. Some adherents of these faiths, in deference to secular orthodoxy, have incorporated the orthodox secular conception of ethics into their faiths, differentiating ethical from
theological concerns, and reserving the term “ethical” for the interpersonal. But others have flatly rejected the secular conception of the scope of ethics. The most striking rejection, perhaps, comes from fundamentalist Islamic thinkers, who reject both the content of non-Islamic secular ethics, as well as claims about morality’s scope. Islam, on this view, is a “total” way of life that governs every single intentional action, from the most “ethically” momentous to the most “ethically” trivial. Islamic precepts therefore govern not just interpersonal life, but personal habits, recreation, and personal belief. We find similar themes in Jewish and Christian writing. Likewise, both ancient Greek and Buddhist ethical systems, though more or less secular, are relatively comprehensive guides to life, not restricted to interpersonal duties.21

We could in principle take these somewhat exotic moral conceptions as counterexamples to the orthodox account of the scope of morality, but we needn’t look so far afield. By the early 1960s, a small but significant handful of writers in the analytic tradition were questioning the idea that morality’s scope is ipso facto restricted to the interpersonal. One of the earliest and most famous examples is P. F. Strawson’s 1961 paper, “Social Morality and Individual Ideal.” In it, Strawson begins by sketching the idea of “an ideal form of life”: “Men make for themselves pictures of ideal forms of life. Such pictures are various and may be seen in sharp opposition to each other; and one and the same individual may be captivated by different and sharply conflicting pictures at different times.”22 As Strawson construed the term, such “pictures” pertain to the overall contours of a life—the basic ends to which the life is devoted, not merely duties owed to others:
The ideas of self-obliterating devotion to duty or to the service of others; of personal honour and magnanimity; of asceticism, contemplation and retreat; of action, dominance, and power; of the cultivation of ‘an exquisite sense of the luxurious’; of simple human solidarity and co-operative endeavour; of a refined complexity of social existence; of a constantly maintained and renewed affinity with natural things—any of these ideas, and a great many others too, may form the core and substance of a personal ideal.\textsuperscript{23}

Note that while these ideals may have social elements—they may be social in content or require a particular social structure for their effectuation—not all of them fit the conception of morality associated with morality. It’s especially difficult to see how either the ideals of ascetic solitude or of “an exquisite sense of the luxurious” could fit that conception: neither ideal makes much sense if conceptualized in terms of other-regarding duties. I can’t have an obligation \textit{to you} to retire to a monastery or to fast, much less an obligation \textit{to you} that I dine at Lutece’s or collect African art.

And yet, Strawson insisted, such ideals are ethical in content: “I think there can be no doubt that what I have been talking about falls within the region of the ethical. I have been talking about evaluations such as can govern choices and decisions which are of the greatest importance to men.”\textsuperscript{24} So the criterion of the “ethical” is not—as it is for morality—other-directed duty, but rather that which can guide action and has “importance.” By this standard, what will count as a morality will far outstrip anything subsumed by morality. Strawson himself mentions that ethical ideals are given “the liveliest” expression in art, and mentions Pascal, Flaubert, Nietzsche, Goethe, Shakespeare, and Tolstoy as paradigmatic examples. With the possible exception of Tolstoy, none of these authors projects an ethical idea captured by morality.
In the same paper, Strawson contrasts personal ethical ideals with what he calls morality. Whether “the ethical…falls within the region of the moral,” he writes, “is something that may be doubted.” The conception of the “moral” alluded to in this contrast is more or less what I’ve been calling morality: “one widely accepted account of the latter [the moral] is in terms of the idea of rules or principles governing human behaviour which apply universally within a community or class”—where the last clause refers, in context, to rules governing the interaction of those in the community or class. This contrast, of course, seems to pit “the ethical” against “the moral,” and much of Strawson’s paper discusses how exactly to reconcile them. In one sense, then, Strawson’s view is a concession to morality: accepting its legitimacy, the question is how to accommodate it to the ethical. In another sense, however, Strawson’s view constitutes a radical departure from morality: traditionally morality was not meant to accommodate (or be accommodated by) anything else; it was overriding. So the very idea that morality might have to make room for a rival concept of “the ethical” undermines the idea that morality has the normative monopoly that it was thought to enjoy.

Strawson’s project points in two possible directions. Either we can try (as Strawson did) to accommodate the ethical to morality, or we can subsume the one into the other, arriving at an altogether new conception of morality that incorporates both. In a little-discussed paper written shortly after Strawson’s, Neil Cooper argues for the latter position—a position that constitutes a more obvious and radical counterexample to morality than Strawson’s view. There is, Cooper argues (echoing Strawson and in some ways prefiguring Frankfurt), a crucial connection between a conception of morality and importance. On this view, “morality in the autonomous sense is not one competing
interest or aspect of life amongst many others. We have to make decisions, which may properly be called ‘moral decisions’, on the order of priority in which we should place these aspects and activities.” Morality’s function is simply to order the priorities: “[A]n autonomous morality provides an answer to the question ‘In what order of importance should our activities, interests, etc. be placed?’ The language of morals in this sense is ‘the language of priorities’.”

Notice that if we accept this view of morality, morality consists of those principles which guide our actions and set our priorities whatever they happen to be, whether self- or other-regarding. On a view of this sort, the idea of an “illegitimate” question—one which subordinates the moral to the non-moral point of view—ceases to make sense.

We can, in light of the preceding, now differentiate moralities on the basis of their scope. On one conception, a morality can consist exclusively of duties to others. Call this a narrow morality. On another conception, morality can consist of duties both to others and oneself. Call this a wide morality.

Pausing to reflect on wide moralities, however, we can distinguish in turn between how duties to self are conceptualized. One relatively narrow conception (of the wide view), deriving in a general way from Kant, restricts duties to self to the traditional roster of self-regarding duties and leaves it at that. Positively speaking, we have duties to cultivate our talents consistently with justice. Negatively, we are bound by prohibitions against self-mutilation, gluttony, sexual over-indulgence, and generally, prohibitions against undermining one’s dignity as an agent. Beyond these duties, however, morality leaves us “breathing room” to do as we please. Our preferences for art and recreation—
our leisure-time pursuits—are not within the province of morality, except insofar as they impinge on some specifically enumerated moral duty. Call this *morality-wide*. A morality-wide includes duties to self, but limits those duties to a circumscribed domain of life, excluding art and leisure-time pursuits.\(^{29}\)

But there is a wider conception (of the wide view) that extends morality’s reach into our leisure time, effectively asserting morality’s scope over all human action. The particular salient feature of this broader conception of morality is the way it governs our aesthetic and (for lack of a better term) “spiritual” lives. Many ethicists and aestheticians have maintained that art is immune to moral evaluation or criticism, but on this conception, art is a moral phenomenon, to be understood in moral terms, and to be created or experienced for moral ends.\(^{30}\) Call this *morality-superwide*. A morality-superwide not only includes duties to self in the traditional sense, but takes such duties to extend to every (or almost every)\(^{31}\) intentional choice, including those bearing on leisure-time pursuits and aesthetic choices.

There is yet another layer of complexity here, derived this time not from ethical concerns but epistemic ones. I have been speaking so far of morality—as most ethicists do—as though morality primarily governed action. But “action” is ambiguous. Because epistemology and ethics are distinct disciplines, it’s tempting to think that epistemologists study belief as distinct from action, while ethicists study action as distinct from belief. But this is an artificial distinction. As I noted in Chapter 1, it is after all *persons* who are justified in believing beliefs, and thereby persons who have or approximate to having knowledge. Likewise, it’s persons who act on the basis of their beliefs. With its interpersonal focus, morality entirely ignores cognitive duties to oneself. But recent
epistemology has begun to focus more closely on cognitive duties to oneself, on virtue epistemology, and on the ethics of belief. So we have yet another conception of morality—a maximally wide morality, one that extends morality beyond leisure pursuits and aesthetic choices to cognitive acts.

5.5 Justification and Explanation

5.5.1 The Exclusivity Thesis

The erotetic dilemma assumes that the interrogative particle “why” is either purely justificatory or purely explanatory. Justification operates in the space of reasons; explanation in the space of causes. But reasons and causes are exclusive: reasons justify, causes don’t. Since we are seeking a justification of morality, “why” cannot be seeking a causal explanation. So it must be seeking a non-causal justification. This assumption entails, as we’ve seen, that the Question is either pointless or illegitimate. Let’s call this the exclusivity thesis, as shorthand for the idea that justification and explanation are exclusive of each other.

The exclusivity thesis may seem so obviously out of touch with the contemporary literature on action theory as to court the danger of coming across as a strawman. But the fact is, it is a stock assumption in discussions of the “Why be moral?” question and has been for decades. So what may seem a strawman to some is evidently an axiom to others. At any rate, in the present context, I’ll sketch four counterexamples to the thesis. The first two—Davidsonianism and consequentialism—are ‘quick and dirty’ counterexamples, not directly relevant to the erotetic dilemma, but subversive of the supposed mutual exclusion of justification and explanation. The second two—agent-causal libertarianism and...
internalism—are somewhat more elaborate, and help us in a more direct way to re-conceptualize the “Why be moral?” question.

5.5.2 First Counterexample: Davidsonianism about Reasons and Causes

One obvious counterexample to the exclusivity thesis is Davidson’s view, advanced in “Actions, Reasons and Causes,” to the effect that reasons are a species of causes, so that “rationalization” (explanation in terms of reasons) is a species of causal explanation. Obviously, if Davidson’s thesis is right, the exclusivity thesis is flatly and obviously false. In that case, Davidson’s thesis would be the only counterexample we needed; we could adopt it without further ado and declare victory against the exclusivity thesis. I’m inclined to think that, suitably qualified, this happens to be true, but the task of formulating the qualifications would take us too far afield for “Davidson adoption” to be the right strategy in the context. We would, among other things, have to deal with the problem of deviant causal chains, weakness of will, freedom of will, and anomalous monism—a clear case of the exegetical tail wagging the conceptual dog. So I mention the Davidsonian counterexample merely to set it aside.

5.5.3 Second Counterexample: Consequentialism

Consequentialism supplies an equally clear if less obvious counterexample to the exclusivity thesis. On a consequentialist view, an action is right if and only if it promotes the best of the options available to the agent at a given time. Notice, however, that “promoting the best available option” is simultaneously a causal and a normative notion:
“promoting” is inherently causal while “best” inherently normative. And both notions are essential to any conception of consequentialism worth the name.

If this is so, “why” questions posed within a consequentialist framework will always have answers that are simultaneously justificatory and explanatory. To ask why $S \phi$’d within consequentialist assumptions is *ipso facto* to ask for a normatively-loaded causal explanation—a causal explanation that, by its very nature, makes irreducible reference to justificatory considerations. For on a consequentialist view, it is true of every action that the agent is either optimizing or not, and so is acting rightly or not. It’s likewise true that rightness will track rightness-producing causes and wrongness will correspond to non-rightness-producing causes. Given the way in which rightness and wrongness track causality in a consequentialist view, and given the consequentialist insistence that the agent always track rightness, causal explanation of an action will always, so to speak, wear its justificatory status on its sleeve.

It might help to put this somewhat formally. Suppose $S \phi$’s at $t_1$ and has effect $E$ at $t_2$. $S$’s $\phi$-ing will be justified insofar as $\phi$-ing at $t_1$ was the best of the available options for producing the best effect at $t_2$. Suppose that $S$’s $\phi$-ing at $t_1$ was perfectly justified. In that case, explanation and justification will coincide: $S$’s action is explained by way of $S$’s selection of the optimal. But now suppose that $S$’s $\phi$-ing at $t_1$ was extremely suboptimal. In that case, explanation and counter-justification will coincide: $S$’s action will be causally explained by way of $S$’s selection of the suboptimal. All action on a consequentialist view will fall into one or the other of these categories. It follows that if
we accept consequentialism, the “why” of justification cannot be wholly distinguished from the why of explanation—another counterexample to the exclusivity thesis.

Unfortunately, while consequentialism would be the perfect counterexample to the exclusivity thesis in other contexts, it cannot serve as one here. For our aim is to formulate a version of the “Why be moral?” question that is neutral as between various ethical views. To assume consequentialism would beg the question against non-consequentialist moralities and so be useless for the task at hand. So I mention the consequentialist counterexample, like the Davidsonian one, merely to set it aside.

5.5.4 Third Counterexample: Agent-Causal Libertarianism

I turn now to the first of two counterexamples that can in fact serve usefully as counterexamples in the present context, of formulating a new version of the “Why be moral?” question.

Let’s return for a moment to the argument for the exclusivity thesis and ask what is supposed to be wrong with an explanatory interpretation of “why.” Suppose that we interpret the “why” to be explanatory. The problem here is supposed to be that a purely explanatory why operates in the space of causes, which threatens the autonomy of reason. Since reasons are what justify (the argument continues), we thus want to avoid a causal interpretation of the “why” in our question. And so (the argument concludes) we’re pushed to adopt a non-causal justificatory interpretation of “why.”

There is a way to circumvent this worry—in one sense simple, in another complex—and that is to adopt agent-causal libertarianism as a metaphysical presupposition of the “why” question. If we do, we kill two birds with one stone.
For one thing, if we do, the worry about “causes threatening the autonomy of reason” evaporates straightaway. On an agent-causal view, the agent is the cause of his choices, and his choices (among them his choice to be rational or not) are entirely up to him. So whatever objections might be made against agent-causal libertarianism, it cannot be objected that its adoption introduces a form of “brute causality” that threatens reason or responsibility.

Second, in adopting agent-causal libertarianism, we introduce an alternative to the purely explanatory and purely justificatory “why.” Since agent causality operates by causality, agent-causal accounts are explanatory; they cite causes. But since agent causality is a form of libertarianism, it leaves the agent free to choose between alternatives. If we assume that the alternatives have some normative ranking, then any contrastive explanation of the agent’s choices—any account that explains why the agent took, say, option 1 over option 2—will have normative significance. A causal explanation that causally explains why Raskolnikov chose to kill the old woman as opposed to not killing her is, on an agent causal account, normatively loaded. If we assume that murder is wrong, the causal account that explains why Raskolnikov chose murder over non-murder will ipso facto also explain why Raskolnikov’s action was wrong. It will, by its nature, pick out those features of Raskolnikov’s deliberations that were under his control and that causally explain why he killed the old woman.

Agent causal libertarianism is of course a controversial metaphysical thesis; I mean neither to deny that nor to defend it here. As I suggested in Chapter 3, arguments for and against metaphysical libertarianism, as well as for and against versions of libertarianism, are methodologically (and epistemically) prior to the foundation of ethics,
and so are methodologically prior to the topics discussed in this study. But a conditionalized claim still drives the point home: if agent-causal libertarianism is defensible, then it constitutes a counterexample to the exclusivity thesis. The antecedent would have to be demonstrated elsewhere, but the conditional is all we need. Even if we’re neutral between libertarianism and non-libertarianism, we have to admit that libertarianism would count as a counterexample to the exclusivity thesis, and admit the counterexample as a possibility. The only way to exclude it would be to reject libertarianism—which would take us from a discussion of meta-ethics to metaphysics, i.e., beyond my current scope.

5.5.5 Fourth Counterexample: Internalism

Let’s return at last to the conclusion of the exclusivity argument, and ex hypothesi (if only for a moment) adopt it as true. If we do, a rather glaring problem arises: in adopting it, we lose any connections between justification, motivation, and explanation. For either (i) the justificatory reason is itself motivational, or (ii) it bears some relation to the agent’s motivations, or (iii) it neither itself motivates nor bears a relation to the agent’s motivations. In the first two cases, there must be some connection between reasons and causality. In case (iii), however, which is the one that the exclusivity thesis entails, the putatively justificatory reason becomes epiphenomenal; it justifies action without affecting the agent, or motivating the agent to do anything. But that is simply nonsensical.

We can avoid the preceding difficulty by espousing some form of internalism. The varieties of internalism are a topic of their own, which it would be counterpurposive
to discuss here.\textsuperscript{36} We need merely adopt a generic form of internalism to see the relevant point. Let internalism be the view that

If an agent \( S \) believes that \( \phi \)-ing is right in circumstance \( C \), then \( S \) will \( \phi \) in \( C \).\textsuperscript{37}

If so, or insofar as this is so, once again, we solve two problems at once.

First, in adopting internalism, we \textit{ipso facto} preclude the possibility that justification will be severed from motivation: internalism preserves an internal connection between the two things; the stronger the form of internalism, the stronger the connection.

Second, internalism gives us an alternative to both a purely justificatory and a purely explanatory conception of reasons for action. Suppose \( S \) \( \phi \)’s in \( C \). Internalism supplies an explanatory scheme for any such action: \( S \)’s \( \phi \)-ing in \( C \) (it says) is causally explained by \( S \)’s belief that it was right to \( \phi \) in \( C \). But if we assume a very minimal form of moral realism, this becomes a normatively loaded explanation: \( S \)’s \( \phi \)-ing is explained by \( S \)’s belief that it was right to \( \phi \) in \( C \), but it is justified by one and the same belief, so long as (or insofar as) the belief is \textit{true}. To put the point another way, on an internalist view, true beliefs about morality simultaneously justify and explain action; false beliefs simultaneously counter-justify and explain it.

Again, internalism like libertarianism is a controversial thesis in the philosophy of action. But what I said about libertarianism holds, I think, of internalism as well. The internalist/externalist dispute is methodologically prior to arguments for and against meta-ethical foundationalism. So once again, the present counterexample has conditional force: if internalism is true, we can reject the exclusivity thesis. But for present purposes,
all we need is a viable alternative to the exclusivity thesis, and even the conditionalized claim is enough to secure that.

5.5.6 Agent-Causal Libertarianism & Internalism & Maximally Wide Morality

Combined

I suggested in the foregoing that metaphysical libertarianism and internalism were both distinct counterexamples to the exclusivity thesis, as they are. But as theses about different aspects of action and agency, they are of course combinable. And they are, in combination, not just consistent but mutually supporting. Libertarianism tells us that the agent is—with respect to some class of actions—free. Internalism tells us that agent is motivated by her moral beliefs. If we combine these two claims with a maximally wide conception of morality, we get the result that agents are free and motivated by their beliefs about what it is right to pursue, where “rightness” is simply a matter of priority and could, as a matter of scope, concern any sort of action.

One way of putting all three of these claims together is to tie them together by way of doxastic voluntarism. An agent can be free with respect to the formation and adoption of her moral beliefs in the maximally wide sense, and be motivated to act on precisely those beliefs (freely adopted) which express her all-things-considered convictions about what it is right to do, where “rightness” ranges over all of the actions in her power across the entire range of her lifespan.

5.6 The Failure of the Erotetic Dilemma and a New Question

We can now easily see why the erotetic dilemma fails. The dilemma only works if it exhausts logical space. But if the arguments of this chapter are right, the dilemma does
not even come close to exhausting logical space. In fact, it confines itself to one very tiny corner of the relevant space. Going exclusively by the conceptions of “should,” “morality,” “I” and “why” canvassed here, we can generate at least one alternative conception of “why” over and above why_e and why_j; several conceptions of “morality” and “should” over and above those dictated by morality_e; and several conceptions of agency unacknowledged and undiscussed by anti-foundationalists. This is of course a somewhat conservative estimate; I have not tried to canvass every possible permutation of each of the relevant concepts.

In one sense, then, the arguments I’ve rehearsed here serve to defeat the erotetic dilemma in a purely quantitative way. I’ve generated more versions of the Question than anti-foundationalists have ever dealt with. But as remarked above, this quantitative point has qualitative implications: the point is not merely that I’ve generated more versions of the Question in strictly numerical terms, but that I’ve shown that the “Why be moral?” question involves a greater range of issues than has previously been thought. The difficulty for the anti-foundationalist is not merely that of churning tediously through various iterations of the Question, but of churning through questions generated by philosophically substantive alternatives opposed to anti-foundationalism. The anti-foundationalist must not only confront a raft of questions, but tackle narrow versus wide moralities, categorical versus conditional conceptions of obligation, libertarian-optimistic versus determinist-pessimistic conceptions of agency, as well as internalism versus externalism. Many anti-foundationalists address some of those issues, but none addresses them all. So the dilemma fails.
The problem for erotetic anti-foundationalism goes further than this. Premise (3) of the erotetic dilemma assumes that the moral point of view and its rejection constitute jointly exhaustive options. Either one affirms the moral point of view or rejects it, but there is no third option. As can be inferred from the foregoing, however, there is a third option. Suppose that one entertains the possibility of a maximally wide conditional morality. Such a morality consists of conditional prescriptions, and includes an ethics of belief as part of its content. If one understands morality in this way, one can ask “Why should I be moral?” and understand that question to be asking the following somewhat hedged question:

If morality includes an ethics of belief, and that ethics of belief includes the evidentialist requirement that I grasp the evidence for any belief \( p \) before I fully endorse \( p \), morality would require me to grasp the rational justification of moral norms before I fully endorsed morality. What \emph{then} is the reason why I should be moral?

Notice that in one sense, this question proceeds from a version of the moral point of view: the purpose of asking the question involves adherence to a moral norm which is not itself put into question in the act of posing the question. In that sense, the question is asked from “within” the moral point of view. But contrary to the claims of erotetic anti-foundationalism, this hardly entails that asking the question is pointless. The moral point of view is precisely what is dictating that the “Why be moral?” question be asked about the \emph{rest} of morality. Notice likewise that “the moral point of view” from which the preceding question proceeds is emphatically not “the moral point of view” associated with morality\(_c\), which is the conception of the moral point of view from which the erotetic dilemma proceeds. In that sense—the one endorsed by morality\(_c\)—the question is asked from “outside” of the moral point of view. But contrary to the claims of erotetic
anti-foundationalism, it begs the question to call the question illegitimate for that reason. It’s not only legitimate, but required by evidentialism. Once again, the dilemma fails.

Having rejected the erotetic dilemma, it’s worth noting that the dilemma does identify a real problem, though not an insuperable one. The problem is that different formulations of the Question can fall into the trap set by the dilemma. A version of the Question that does tacitly assume the legitimacy of the moral point of view (in the sense implied by morality) does involve a certain pointlessness about it. It’s very easy to demand a justification of morality while assuming its legitimacy, and then to remain unsatisfied at the fact that no justification conveys novel information. And no justification could convey novel information if the information has already been packed into the request for justification. By the same token, a version of the question that simply assumes the perspective of immorality and then demands that morality accommodate itself to that perspective really is illegitimate. Apart from the psychological difficulty of convincing the potentially unconvincible, there is the added difficulty that if morality is never a means to immorality, the very form of the Question blocks the truth about what morality really is.

In admitting these facts, we need not concede the truth of the erotetic dilemma as stated. As stated—or so I’ve argued—the dilemma fails. What we should admit is the grain of truth in the dilemma, despite its failure. A successful version of the “Why be moral?” question must at a minimum avoid these twin difficulties. For one thing, it can’t make far-reaching moral assumptions of its own, or start with controversial assumptions about the nature of morality. If it does, it will not really pose a real question, and so fall into the “pointlessness” problem. Second, it cannot adopt an immoral or anti-moral
posture (as depicted by the stock figures in the literature—Plato’s Thrasymachus and Callicles, Hobbes’s Foole, Hume’s Knave, etc.). If it does, it may well fall into the “illegitimacy” problem. But third, it has to ask a real question that seeks a real answer. A question that merely avoids the first two difficulties will not necessarily ask a question worth asking.

The place to begin, I think, is with the combination of claims asserted in section 5.5.6 above: agent-causal libertarianism, internalism, and a maximally wide morality. Suppose we adopt these three claims and imagine an agent \( S \) of whom we predicate them. In this case we know that \( S \) has the power to act on alternatives and that the reasons for and causes of her actions coincide in the sense described in section 5.5.6. \( S \) faces alternatives and as free, can act on any of the alternatives within a set. To act, \( S \) weights alternatives, choosing between the weights, where the weights are themselves determined by her conception of what she regards as right.

To get from that account of agency to our Question, we need only ask the following: for any alternative (or proper subset of alternatives) that \( S \) faces, what is the explanation (compatible with libertarianism) for why \( S \) has to choose one alternative over another? In virtue of what, in other words, does one possible option get selected over another one? Why does it make a difference which of a range of alternatives \( S \) selects or prioritizes when she faces a range of them? To put the same question in the first-person: If I am a free agent, and face options \( \{O_1 \ldots O_n\} \) at \( t_1 \), what fact explains why and in what sense I have to select any of those options, and if so, rank some options over others? Notice that this question makes no assumptions at all about the content of morality (except that the Question needs an answer). It doesn’t even assume that I must select an
option. It merely points out that I could select an option or not, and if I do, must select one option over others. *What makes it the case that I must (ever) rank some options over others, given that I’m free to select any option from the set of options I face, and indeed, could in principle freely give up on the very idea of selecting anything?*38 This question, as I see it, entirely evades the erotetic dilemma, and offers a powerful motivation for a meta-ethical foundationalism.

First, if I ask the question sincerely, it’s not really clear whether I’m “within” or “outside” of the moral point of view. In one sense, I’m outside of it because I’m not adopting the moral point of view as understood by morality. In another sense, if I regard the asking of the question as the expression of a desire to understand the basis of morality—and if I take this desire to be justified in a merely *prima facie* way—that fact moralizes the question. And so premise (3) of the dilemma is (once again) falsified.

Suppose, however, that we take the preceding to be evidence that I’m inside the moral point of view. Does that mean that the question is pointless? No. I could give a *prima facie* presumption to the idea that I have an obligation to understand the source of moral norms while putting the rest of my moral obligations fully in doubt, and asking the question from that perspective. From that perspective, the question has a distinct goal and seeks novel information. Hence it is not pointless, and premise (1) of the dilemma is (once again) falsified.

Suppose now that a critic insists that in doubting morality so severely, I have put myself outside of the moral point of view. Is the question problematically explanationist in the sense discussed in Chapter 4? No, because while it seeks a causal explanation, the sort of explanation it seeks—an agent-causal explanation—is not problematic in the way
that explanationism is.\textsuperscript{39} Is the question subordinationist? No. As far as I can see, the
question is entirely neutral on the form and content of morality. Nor does it assume the
legitimacy of an amoral, anti-moral, or immoral perspective to which it demands
morality’s subordination or conformity. That falsifies premise (2), and thereby scuttles
the dilemma.

I conclude, then, that the erotetic dilemma is a conclusive failure. Though there
may be other objections to meta-ethical foundationalism, it cannot be said that the
question itself is illegitimate. And so, we’re entitled to ask it, to offer an answer, and to
take the matter from there—as I do in the next chapter.\textsuperscript{40}

5.7 Notes

\textsuperscript{1} John Rawls, \textit{A Theory of Justice}, rev. ed. (Cambridge: Belknap Press, 1999),
sec. 19.

\textsuperscript{2} Rawls, \textit{Theory of Justice}, pp. 98 and 99 passim.

\textsuperscript{3} Philippa Foot, “Morality as a System of Hypothetical Imperatives”
\textit{Philosophical Review}, vol. 81, no. 3 (July 1972), pp. 305-16, is the exception (or perhaps
one of the exceptions) that proves the rule.

\textsuperscript{4} Stephen Darwall, \textit{Impartial Reason} (Ithaca, NY: Cornell University Press,

\textsuperscript{5} Alasdair MacIntyre offers an apt critique of this sort of argument in the Preface
of \textit{After Virtue}: “The notion that the moral philosopher can study \textit{the} concepts of morality
merely by reflecting, Oxford armchair style, on what he or she and those around him or
her say is barren. This conviction I have found no good reason to abandon; and
emigration to the United States has taught me that when the armchair is in Cambridge,
Massachusetts, or in Princeton, New Jersey, it functions no better” (MacIntyre’s italics;
Suppose that $S$ holds his breath and commits an immorality the commission of which kills him, hence stops his breathing? In that case, a proponent of the morality-conditional-on-breathing view would have to rely on the first of the two responses in the text.

A final counterexample: suppose $S$ initiated an (otherwise) immoral action without breathing, then effected the action without breathing, never to take a subsequent breath? Would that action be beyond moral judgment? The answer is “yes,” but the “yes” is not much of a concession. For one thing, the counterexample is outlandish to the point of raising questions about its sheer possibility. But even if the counterexample were possible, what it would show at best was the implausibility of the morality-conditional-on-breathing view, not morality’s conditionality per se.

I say “loosely modeled” because Gewirth in fact conceives of morality as categorical in binding force. But we could, mutatis mutandis, imagine a view like that described in the text, Gewirthian in all respects except categoricity.

I’ve omitted two arguments from the text because they strike me as fundamentally different in character from the ones described in the text, and ultimately idiosyncratic.

One type of argument insists that morality and self-interest are just axiomatically distinct sources of practical norms. In one version, this claim is simply insisted on as a sort of intuitively obvious proposition not in need of further argument or explication; in that version, it obviously begs the question, and nothing need be said about it but that. It should be “intuitively obvious” that an advocate of an egoistic morality is not going to concede his position simply because someone finds it counter-intuitive. Another version holds that moral and egoistic norms are just too different from one another to be grouped in the same category. But this is just another way of begging the question: an egoist won’t think them different; he’ll think them identical. Another version of the argument holds that a self-regarding morality is too trivial in content. But since the egoist thinks that all of morality is self-regarding, this objection is an ignoratio elenchi. Yet another “argument” (not typically aimed at egoism, but in principle applicable to it) asserts that a self-regarding morality is too “priggish.” Insofar as this is an ad hominem, it’s fallacious; insofar as it’s a complaint, it begs the question.

A second type of argument, associated with Parfit, holds that egoism is self-defeating (Derek Parfit, *Reasons and Persons* [New York: Oxford, 1987], pp. 3-12). The argument is that egoism assigns to us the aim of trying to make our lives go as well as possible, but it can (according to Parfit) be the case that cultivating the disposition to make my life go as well as possible will make it go worse. The argument is easily answered: Suppose we adopt an Objective List theory according to which cultivating a disposition to make one’s life go as well as possible will make it go better. In that case, the state of affairs Parfit describes—in
which cultivating the disposition to make one’s life go well undercuts that aim—will
simply be impossible, and his objection falls immediately to the ground. The sheer
obviousness of this response to Parfit coupled with his failure to deal with it suggests that
Parfit’s account of self-interest suffers from a version of the difficulty described in the
text. In any case, Parfit himself concedes that self-interest doesn’t fail on its own terms
(pp. 11-12), so it’s hard to see how his argument constitutes an objection to a self-interest
theory.

E.g., Brian Medlin, “Ultimate Principles and Ethical Egoism,” Australasian
1999), ch. 6; James Rachels, “Egoism and Moral Scepticism,” reprinted in Vice and
Virtue in Everyday Life: Introductory Readings in Ethics, ed. Christina Hoff Sommers

Kurt Baier, “Egoism,” in A Companion to Ethics, ed. Peter Singer (Malden,

Laurence Thomas, “Ethical Egoism and Psychological Dispositions,” American

G. E. Moore, Principia Ethica, rev. ed., ed. Thomas Baldwin (New York:

William James, “The Consciousness of Self,” in Self-Interest: An Anthology of
My discussion of this topic, as well as the ascription of the view to James, owes a debt to
a lecture by Kelly Rogers, “Ethical Egoism in Historical Perspective,” University of
Virginia, 1997.

One notable and very rare exception is J. L. Mackie’s discussion in J. L.
while Mackie makes his assumptions about the source of self-interest explicit, he offers
no defense of them beyond citations to the classic texts in which they’re discussed. One
source is Hobbes’s Leviathan, which describes us as inclined toward predation given our
propensity to compete over scarce resources. Mackie infers that these deep features of
human nature determine our interests, and so would make untrammeled self-interest a
predatory or malevolent phenomenon. But as Mackie offers nothing in the way of
argument for either of Hobbes’s claims, we are, I think, justified in dismissing them
without further ado. It’s worth noting that Mackie’s discussion is by far the most self-
conscious I have seen in the literature.

For a good discussion of this issue, see Stephen R. C. Hicks, “Ayn Rand and
Contemporary Business Ethics,” Journal of Accounting, Ethics and Public Policy, vol. 3,
15 Cf. Laurence Thomas, “Ethical Egoism and Psychological Dispositions.” Thomas’s argument strikes me as a nearly perfect example of *ignoratio elenchi*.


17 Here I follow Mackie’s very clear discussion of universalizability; see Mackie, *Ethics*, ch. 4.

18 There probably are some narrow competitive or adversarial contexts in which it would be egoistically irrational to identify the advantage of one’s competitor or enemy and recommend it. Here we need to distinguish firmly between competitors and enemies, however. A fellow competitor is someone with whom one (in principle) shares a moral outlook but with whom one happens to be competing. It may not be in one’s interest to identify one’s competitor’s advantage to him, but the competition takes place within a backdrop of rules which it is in both competitors’ interest to follow. An enemy, by contrast, is someone who is hostile to one’s moral outlook. Egoism may well be incompatible with offering comfort to enemies, but this is hardly unique to egoism, and not clearly an objection to it.


26 The view prefigures Frankfurt’s notion of “the importance of what we care about,” as in his essay collection of that name. But while “importance” plays an important role in Frankfurt’s thought, his official view about morality echoes Strawson’s: morality is a set of interpersonal duties.

27 Neil Cooper, “Morality and Importance,” in The Definition of Morality, both quotations from p. 97.


29 A view of this sort could in principle be consistent with the idea that some choices were morally indifferent or optional, but moral principles would themselves explain which ones were. But it is also consistent with the possibility that, strictly speaking, no choices are morally indifferent: it could be that moral principles tell us which choices are trivial and hence a matter of pure preference, but the same principles would not only identify which choices fit this description, but might well insist that we cultivate strong preferences (in this context) so as to resolve what would otherwise be dilemmatic situations where no principle decided between the options. Thanks to Consuelo Preti, Hilary Persky, Carrie-Ann Biondi, and Allan Gotthelf for helpful discussion on this issue.


31 See note 29 above for an explanation of the qualification.

33 For an interesting example of such a maximally wide morality, see Henry David Thoreau, “Where I Lived, What I Lived For,” in Thoreau, *Walden*: “I know of no more encouraging fact than the unquestionable ability of man to elevate his life by a conscious endeavor. ...To affect the quality of the day is the highest of arts. Every man is tasked to make his life, even in its details, worthy of the contemplation of his most elevated and critical hour” (in *Walden and Other Writings*, ed. Joseph Wood Krutch [New York: Bantam Books, 1962], p. 172). Thanks to Carrie-Ann Biondi for drawing my attention to this passage.


37 Here I adopt and modify what Smith (citing Nagel, McDowell, and Platts) calls strong internalism; Smith, *The Moral Problem*, p. 61.

David Solomon points out that my definition of “internalism” is both idiosyncratic and stronger than most. Typically, internalism is defined as a view according to which S’s judgment that phi-ing is right motivates a defeasible tendency toward phi-ing. I define it instead so that, given relevant circumstances, S's judgment that phi-ing is right makes S phi.

In one sense, we might well regard these definitions as effectively equivalent. Whereas the standard definition takes moral judgments to be defeasibly motivating, my definition takes moral judgments to move the agent all the way to action in specific circumstances but not others. We might take the defeasibility criterion of the standard definition to be equivalent to the excluded circumstances in my definition: internalism will be defeated on the standard view by those circumstances I exclude.

I do think that my definition is somewhat clearer than the standard one, however. If the relation between S’s judgment and S’s action is defeasible (as in the weak internalist definition), then it is possible that S always judges that phi-ing is right, and yet never
phi’s—namely, in the case where phi-ing always happens to be defeated. On my view, by contrast, once we exclude the cases where internalism was never meant to apply—cases where what prevents the judgment from producing the action is external to the agent—the judgment directly produces the action, an implication that seems truer to the spirit of internalism than the thought that a judgment tends to motivate action, but could still be defeated by something else.

Someone might object that the “stronger” definition still yields the same result as the weaker one, since it is possible that S might always judge phi-ing to be right and still never phi, because he always happened to be in precisely those circumstances excluded by the strong definition. But we might (as I think we should) define the excluded circumstances right from the start so that most agents would be in the included circumstances most of the time. On this view, the circumstances in which internalism does not operate are known right from the outset; once we put these aside, strong internalism asserts an indefeasible connection between judgment and action. And if there turn out to be few such circumstances, strong internalism will end up being a very strong thesis, obviously incompatible with any form of externalism. This contrasts with a view, like weak internalism, that specifies no indefeasible connection between judgment and action. Since we don’t know from the outset what will defeat internalism, a view like this could in principle collapse into, or be compatible with, a weak form of externalism.

I thank Carrie-Ann Biondi, Jason Raibley, and Michael Young for hours of discussion on this topic.

38 I take this question to be conceptually equivalent to the version of the question cited in the introduction of the chapter.

39 What I mean to say is that it is not reductionist in the way that explanationism is, but having used “reductionism” in a completely different sense in the Introduction, it would be confusing to use that term here.

40 Thanks to Carrie-Ann Biondi, Allan Gotthelf, Daniel Jacobson, Rick Kamber, Jason Raibley, Consuelo Preti, David Solomon, Michael Young, and an audience at The College of New Jersey for helpful comments and conversation on the subject of this chapter. A special thanks to Carrie-Ann Biondi for editorial help.
6.1 Introduction

As I suggested in the Introduction, the “foundations of ethics” is an ambiguous phrase intended to describe a somewhat equivocal problem. In one sense, the “foundations of ethics” seems to denote an epistemic problem: how do foundationalist epistemic strictures apply to moral belief? In another sense, the same phrase seems to denote not an epistemic but a meta-ethical problem: what sort of theory (and ultimately, what theory) successfully answers the “Why be moral?” question? Given this, epistemic foundationalism and meta-ethical foundationalism may be seen as homonymous, analogous, reductionist, or simultaneously a reductionist-analogy hybrid. A discussion of the foundations of ethics needs both to distinguish the two sorts of inquiry and to show the relations between them. Ideally, my argument suggests, an inquiry into the foundations of ethics should yield a foundationalist moral epistemology, a foundationalist meta-ethics, and an account, consistent with both, of how the epistemology coheres with the meta-ethics. The result will be a comprehensive normative foundationalism: a single foundationalist theory consisting of two distinct but related sub-theories.
In this chapter, I want—taking the preceding claims as backdrop—to sketch my own positive account of such a theory. The account has three parts. (1) In Chapter 2, I argued for the possibility of a non-intuitionist foundationalist moral epistemology which I called Aristotelian empiricism. In section 6.2 below, I describe the methodological implications of Aristotelian empiricism by way of a contrast with the method of reflective equilibrium. (2) In Chapter 5, I established the legitimacy of meta-ethical foundationalism, formulating a version of the “Why be moral?” question intended to motivate that view. In section 6.3, drawing on a principle I call “the conditionality of life” I argue for an ethical egoist answer to that question.¹ (3) Finally, in section 6.3, I describe the structure of a comprehensive normative foundationalism consisting of the conjunction of Aristotelian empiricism and ethical egoism. The structure turns out to instantiate what in the Introduction I called the “epistemic reductionist-analogy hybrid.”

6.2 Epistemology and Method: Aristotelian Empiricism versus Reflective Equilibrium

As I argued in Chapter 2, Aristotelian empiricism is a form of epistemic foundationalism based on direct realism and a theory of abstraction. Epistemic foundationalism holds that epistemically justified cognition has a two-tiered structure, foundational and superstructural. The Aristotelian empiricist version of this view holds that the foundation of justified cognition consists of sensory perceptions as construed on a direct realist theory of perception, which are abstracted into concepts and eventually into propositions to which we give or withhold doxastic assent. Beliefs are justified insofar as we can trace them back, via abstraction, to their origins in the evidence of the senses. The basic task of an empiricist moral epistemology would therefore be to trace
our moral concepts in this way to the evidence of the senses—to give an account, in other words, of how our moral concepts are formed \textit{ab initio}.

As I suggested in Chapter 2, the empiricist project so conceived is not a stand-alone endeavor; it has to work in tandem with a separate account of what all of our (legitimate) ethical concepts have in common, which is the task of a meta-ethical foundationalism. I cannot, in the present context, lay out a full account of the empiricist project—a full account of the nature of ethical concepts. What I’ll do instead is to indicate (some of) the methodological implications for ethical inquiry of espousing an Aristotelian empiricist conception in moral epistemology. As it turns out, these implications are both significant and radical, and are best seen by contrasting the methodological implications of Aristotelian empiricism with those of reflective equilibrium.

Dale Jamieson describes the method of reflective equilibrium (MRE) as follows:

These days some version of coherentism is the dominant view of what constitutes proper method for theory construction in ethics. Coherentism can be roughly characterized as the view that beliefs can be justified only by their relation to other beliefs.

The most influential form of coherentism is Rawls’s method of reflective equilibrium. According to Rawls, proper method involves beginning with a set of considered beliefs, formulating general principles to account for them, and then revising both principles and beliefs in the light of each other, until an equilibrium is reached.²

This description of MRE prompts a number of objections. The objections, as I see them, are fatal both separately and in conjunction. Some are unique to coherentist theories, others to reflective equilibrium, and together they help to explain the advantages of
foundationalism generally, and of Aristotelian empiricism in particular, over reflective equilibrium.³

1. MRE begs the question. A first problem is that insofar as MRE is coherentist, and coherentism countenances (or requires) circular reasoning, any theory based on reflective equilibrium will be (avowedly) circular. But if circularity is a fallacy, then all such theories will be fallacious, and no fallacious theory can count as justified or as knowledge.

Though some might contest the claim that circularity is always fallacious, in the present case it seems obvious the MRE enjoins a form of circularity that begs the question. We begin, we are told, with a set of “considered beliefs,” and then “account” for these beliefs by appealing to higher-order principles. There is a two-fold assumption embedded into the method, each one of which begs the question. The first is that it is clear what counts as one of “our considered beliefs.” The second is that “our considered beliefs,” whatever they turn out to be, are probative simply because they are ours. The first assumption overlooks the possibility, discussed at length by MacIntyre among others, that “our” beliefs are simply too fragmented and disorganized to have sufficient unity to be discussed in one breath.⁴ The second assumption overlooks the possibility that the “considered beliefs” in question are simply too far off base to constitute the proper starting point for inquiry. As far as I can tell, no defender of MRE has successfully answered either criticism. Indeed, it is telling that defenders of MRE habitually beg the question against both criticisms.⁵

2. MRE is methodologically vacuous. Even if we waive the first objection, a second one arises, namely, that while MRE purports to be a method, it is
methodologically indeterminate. MRE tells us to achieve a coherentist “equilibrium” between our theoretical and pre-theoretical beliefs. But apart from the fact that the method gives us no account of what counts as one of “our beliefs,” it gives us (a) no determinate account of the criteria for “equilibrium,” and (b) no determinate advice for achieving equilibrium, however that concept is understood.

Consider (a). The “equilibrium” to which MRE alludes is, obviously, a species of coherence. Unfortunately, it’s unclear what species of coherence it is, because it’s unclear what the genus “coherence” denotes in the first place. On the weakest interpretation, a coherent belief-set is one that is free of contradiction; on a stronger interpretation, every belief in such a belief-set entails every other. The difference between the two interpretations is stark, but the method gives us no way of discriminating between them. Since it doesn’t, it neither gives us a determinate goal at which to aim nor determinate prescriptions by which to achieve its aim. A fortiori, it gives us no goal or methods conducive to justified moral belief or moral truth.

A defender of MRE might claim that the conception of “coherence” in question is explanatory: higher-order principles have to explain, or be the best explanation of, our considered beliefs. But the conception of explanation involved here is entirely obscure. If all we need is an explanation, then it’s unclear how MRE decides between competing explanatory hypotheses (whatever counts as one). But if what MRE requires is the best explanation, it is likewise unclear what the criteria for “bestness” are supposed to be.6

Now consider (b). MRE tells us to fit our “pre-theoretical” to our “theoretical” beliefs. The distinction between the two sorts of belief is obscure enough, but since the method offers no weighting of the one sort of belief to the other, the prescription turns
out to be meaningless in any case: two agents starting with identical belief-sets could well follow MRE but arrive at diametrically opposed belief-sets by following radically different procedures consistent with MRE. Imagine that Abdul and Bilal begin with the same beliefs, both adopting MRE. Abdul brings his beliefs into equilibrium by taking a highly anti-theoretical strategy, keeping only his pre-theoretical beliefs and driving out all of his theoretical ones. Bilal does the reverse. By the terms of MRE, neither Abdul nor Bilal has done anything wrong—even though each has done the reverse of what the other has done, and each has come to the reverse conclusions as those the other has reached. This result suggests that MRE is less a “method” of ethical theorizing than the absence of one. For while we could describe both Abdul and Bilal as “following the method of reflective equilibrium,” we could with equal accuracy and greater simplicity describe them both as merely following their whims. A method this contentless hardly deserves the name.

3. MRE is incompatible with moral realism or moral knowledge. Finally, even if we take the most charitable attitude toward MRE and waive the first two problems entirely, we’re still left with the fact that the method aims at no more than coherence. But coherence is a relation between beliefs, while knowledge is a relation between beliefs and a world that exists independently of beliefs. If so, the professed goal of the theory is irrelevant at best (and orthogonal at worst) to truth as the goal of cognition, a truth-conducive conception of epistemic justification, and by implication, any form of moral realism. After all, if we begin with nothing but considered beliefs that are false, and merely put them into reflective equilibrium, we still end up with false beliefs, however elegantly expressed. Indeed, the very elegance of their expression might divert us even
father from moral knowledge than we might otherwise have been, since false beliefs elegantly presented are likely to attract us more than false beliefs without such theoretical embellishments.

By contrast with MRE, Aristotelian empiricism requires that we identify, by way of (some version of) meta-ethical foundationalism, a meta-ethical foundation for ethics by way of inquiry into (some version of) the “Why be moral?” question. In answering that question, this foundation—Kantian, utilitarian, egoist, whatever—simultaneously tells us what all ethical concepts have in common. Since our claims will be epistemically justified insofar as we can trace those concepts back to the evidence of the senses, Aristotelian empiricism has the methodological implication that we do what we can to trace them back in that way.

We might describe the method of Aristotelian empiricism (MAE), then, in summary form as follows: (1) Identify the answer to the “Why be moral?” question, (2) derive ethical claims from the meta-ethical foundation identified in (1), (3) use the meta-ethical foundation to identify what all ethical concepts have in common, and (4) trace back the constituent concepts of ethical claims to their basis in the evidence of the senses. This is, to be sure, a demanding set of methodological strictures. But demanding as it may be, it has none of the liabilities of MRE.

For one thing, MAE cannot be described as circular. As a species of foundationalist theory, MAE is a specific answer to the epistemic regress problem. But qua foundationalist answer to that problem, it is explicitly non-circular; if it were circular, it could not by definition be foundationalist. And so MAE avoids the first problem to which MRE falls prey.
Nor can MAE be described as methodologically indeterminate. If anything, just the reverse is true: it gives us four substantive methodological tasks to undertake, and an order in which to undertake them. Moreover, it clearly avoids the Abdul/Bilal problem I sketched above: Abdul and Bilal cannot, consistently with MAE, start at identical places and end at different ones. They both have to answer the “Why be moral” question, and since the various answers to that question are incompatible, there can only be one right answer. They both have to derive norms from this foundation, but some things will count as good derivations and other things won’t. They’ll both have to identify, on the basis of the foundation, what ethical concepts have in common, but they cannot consistently with the theory reach incompatible answers to that question. And they’ll have to trace their ethical concepts back to the evidence of the senses, but while there may be variation within a range, the account they give will have to show how the relevant concepts were formed by abstraction from relevant perceptual similarities. There will be limits on the ways in which that can be done.

Finally, MAE has much less of a problem than does MRE with respect to truth and knowledge. The problem for MRE with respect to truth and knowledge is its coherentism—the so-called “input problem.” But MAE has no such problem. Since the foundation of justified belief on MAE is a direct realist theory of perception, MAE *ipso facto* involves input from the world. If we define the principles of derivation from foundation to superstructure as truth-conducive, we will, to be sure, confront the large task of identifying which principles really are truth-conducive. But there will be no problem in principle—as there is with MRE—of bridging the gap between beliefs on the one hand and the world on the other.

6.3.1 Internalism, Libertarianism, and Meta-Ethical Foundationalism

In Chapter 3, I laid out the basic idea of a meta-ethical foundationalism: a non-epistemic approach to right action according to which foundational items are ethically basic norms that asymmetrically justify a superstructure of non-basic norms. In Chapter 4, I laid out the fundamental objection to meta-ethical foundationalism, and in Chapter 5, I answered the latter objection, vindicating meta-ethical foundationalism and reformulating the question that gives rise to it. The question, recall, was: “What is the justification, compatible with internalism and metaphysical libertarianism, for why a rational agent must rank alternative courses of action and act on the rankings that he or she sets?” In this section, I sketch an egoistic answer to that question based on what I’ll call the conditionality of life.

Internalism is the thesis that if $S$ judges that $\phi$-ing is right, $S$ will $\phi$. Libertarianism is the thesis that for some values of $\phi$, it is the case that $S$ can either $\phi$ or $\sim\phi$, and if $S$ $\phi$’s at $t$, $S$ could have $\sim\phi$’d at $t$, i.e., done otherwise than he actually did. Suppose we stipulate for present purposes that libertarianism applies to any value of $\phi$ where the choice between $\phi$-ing and $\sim\phi$-ing is a choice between rightful and wrongful action. In that case, we face a moderate puzzle. Internalism tells us that if $S$ judges that $\phi$-ing is right, $S$ will $\phi$. But libertarianism tells us that in facing the choice whether to $\phi$ or $\sim\phi$, it’s up to $S$ which of the two he’ll take. Clearly, to make the two claims compatible, we have to say that
once \( S \) judges that \( \phi \)-ing is right, \( S \) eliminates the alternative of \( \sim \phi \)-ing,’ and \( \phi \)’s straightaway; but whether or not to judge \( \phi \)-ing right is up to \( S \).

Internalism, then, can be made compatible with libertarianism, but in making them compatible, we create a new, and larger puzzle. Given internalism, we know that once \( S \) judges an action to be right, \( S \) will take the action. We also know that \( S \) is free whether or not to judge a given action to be right. For that matter, \( S \) is free whether or not to judge it right to act at all. The question we face, then, is this: \( S \) is metaphysically free to form evaluative judgments or not, and thereby metaphysically free to act or not. If libertarianism is the case, nothing can literally force \( S \) to judge anything worthwhile, or to engage in any particular action. In making the judgment that some acts are right, \( S \) is thereby committing himself voluntarily to eliminating what would otherwise be open alternatives to action. Given libertarianism, before \( S \) judges that \( \phi \)-ing is right, \( S \) is capable of either \( \phi \)-ing or \( \sim \phi \)-ing. Given internalism, once \( S \) judges that (say) \( \phi \)-ing is right, \( S \) is eliminating his own capacity to \( \sim \phi \). Our question, then, can be rephrased as follows: What requires \( S \) to do this—and requires it of him on an ongoing basis? In other words, what makes it the case, consistent with metaphysical libertarianism, that \( S \) must voluntarily choose to put specific limits on his own options in a principled way across the span of his lifetime?

One answer developed in the Aristotelian tradition, and given a particularly powerful recent formulation by Ayn Rand, turns on what may be called the conditional character of life as a teleological phenomenon. Rand’s version of the principle involves a broadly meta-biological claim applicable to all (and only) organisms as such, and a
narrower application of that principle to the moral case, intended to answer a version of the preceding question. The answer will turn out to be that $S$ is obliged voluntarily to impose limits on his freedom by adopting goals he endorses as right, and is obliged to do this because his very existence and identity qua human depends on his doing so. Moral agents can only maintain their existence and identity by judging some actions as conducive to their survival and well-being, and others as vitiatory of it. In equating the former with the right and the latter with the wrong, and adopting goals accordingly, we get an account of moral obligation consistent with internalism and metaphysical freedom: there is a sense in which a free agent must in a principled way endorse some ends and reject others, if he is to maintain his existence and identity qua human.

This is of course an extremely bold and abstract claim, and in the present context I can only offer a sketch of it. In what follows, I’ll begin with the broader meta-biological version of the principle applicable to organisms as such, explaining in a general way what’s meant by the conditional character of life (6.3.2). I’ll then apply the thesis to the human case, explaining its application to specifically moral value (6.3.3). From there, I’ll show how the principle gives rise to particular moral prescriptions (in the form of a virtue ethics), and offer some remarks on its egoistic character (6.3.4-6.3.5).

6.3.2 Laying the Foundation (1): The Conditionality of Life Applied to Organisms as Such

Rand states the meta-biological version of the conditionality thesis in the following dense passage:
There is only one fundamental alternative in the universe: existence or non-existence—and it pertains to a single class of entities: to living organisms. The existence of inanimate matter is unconditional, the existence of life is not: it depends on a specific course of action. Matter is indestructible, it changes its forms, but it cannot cease to exist. It is only a living organism that faces a constant alternative: the issue of life or death. Life is a process of self-sustaining and self-generated action. If an organism fails in that action, it dies; its chemical elements remain, but its life goes out of existence.

Life can be kept in existence only by a constant process of self-sustaining action. The goal of that action, the ultimate value which, to be kept, must be gained through its every moment, is the organism’s life.

An ultimate value is that final goal or end to which all lesser goals are the means—and it sets the standard by which all lesser goals are evaluated. An organism’s life is its standard of value: that which furthers its life is the good, that which threatens it is the evil.

The basic idea is this: Living things are ontologically unique entities, and are distinguished from non-living things by virtue of the fact that their existence and identity depend on their engaging in a complex repertoire of continuous action across a lifespan, generated and directed by the organism itself. The nature or identity of an organism determines the kind of action appropriate to it. Precisely because each organism has a determinate nature, each has specific needs, and because it has such needs, it has a determinate mode of life appropriate to its capacities and environment. Given the determinacy of modes of life so conceived, there is such a thing as acting in accordance with, and acting against, the requirements of the organism’s life—the most obvious requirements being those of physical health and functionality, the less obvious including proper mental functioning and psychological health (in the relevant organisms) as determined by the organism’s traits of character. An organism’s default on its mode of life undermines or impairs its functionality by degrees; severe default undermines its
existence and identity altogether and leads ultimately to death. This latter fact partitions
the world for that entity into normatively salient categories: the beneficial, the harmful,
and the indifferent.

This thumbnail sketch of the conditionality thesis involves two important
subclaims, each implicit in the sketch but crucial to the overall intelligibility of the thesis.
There is first a claim of essentialism: every organism has an essence, consisting of the
property or set of properties which asymmetrically explain how the organism engages in
valuation, i.e., how it preserves itself. In conscious organisms, this will typically be the
type of consciousness it has: how it perceives the environment, and how it motivates and
coordinates goal-directed action.

There is secondly a claim about lifespan interests: every organism will pursue a
repertoire of goals that retains its identity across the whole of the organism’s lifespan,
and in so doing, preserves the organism’s existence and identity across that lifespan.
Nutrition is a paradigm case. A biology textbook defines nutrition as “[a]ny process
[autotrophic or heterotrophic] whereby an organism obtains from its environment the
energy and atoms required for maintenance, growth, reproduction, etc.”¹⁰ Somewhat
more formally, an act counts as nutritive insofar as it (i) plays the just-specified
physiological role, (ii) plays that role over the full lifespan of the relevant organisms, and
(iii) its playing the role mentioned in (i) maintains the existence over a full lifespan of
organisms of that kind. Though clause (iii) is obvious in the case of nutrition, it has less
obvious applications, in biology as in ethics. Thus a trait could count as a lifespan
interest on the same pattern as nutrition, insofar as it played a certain psycho-physical
role over the course of an organism’s life, where its playing that role maintained the existence and identity of the organism across the organism’s lifespan.

The latter two claims—about essences and life-span interests—are related in an important way. An organism’s preservation of its essence will itself be its fundamental lifespan interest. The conditionality thesis tells us that an organism must act in specific ways in order to preserve its existence and identity across its full lifespan. It tells us, also, that every organism possesses a set of capacities that enable it to do just that.

The organism’s actualization of those capacities—its essence—is thus both means and end. In one sense, the actualization of the essence enables the organism to act, and so enables it to survive in the obvious sense of enabling it to pursue the repertoire of lifespan interests that will maintain its life. And so preservation of an organism’s essence is a means to its survival. In calling an organism’s lifespan interests a “repertoire,” I mean to convey the idea that the valuation in question will typically be complex, structured, and of some significant duration relative to the organism’s lifespan. Hunting is a means of preserving (say) a predator’s life, but is also a complex structured part of it—as is, say, foraging in the case of herbivores, play in the case of dolphins, trade in the human case, and so on. The organism’s essence makes it possible for organisms of that type to engage in the repertoire of actions instrumental to self-preservation. But in so doing, the organism’s essence constitutes the life and self that is thereby to be preserved. And so preservation of an organism’s essence constitutes the identity of the organism, and takes pride of place among the values it seeks.

Let’s call the complex fact described in the preceding paragraphs of this section the conditionality of life, the thesis expressing it the conditionality thesis, and the entities
to which it applies, *conditional entities*. I’ve suggested that the conditionality thesis somehow answers the “Why be moral?” question, but that of course requires elaboration. To see its relevance, we need to note several implications of the thesis which, though true, are inessential to the “Why be moral?” question. The irrelevance of the irrelevant implications should help to bring out the relevance of the relevant ones.

One clear implication of the conditionality thesis is that life is a necessary condition of valuation: if all and only living things engage in valuation, then if *x* is to value, *x* must be alive. Non-living things—rocks, rivers, planets, stars, gases, minerals—cannot be said to act or value, and so, nothing is valuable to or for them. But though true and implied by the conditionality thesis, life’s being a necessary condition of valuation is not the central idea behind it. For one thing, it’s relatively trivial. For another, it has no important bearing at all on our initial question. The necessary condition claim under consideration tells us that *if x is to value, then x has to be alive*; it cannot be dead and it cannot be inanimate. While plausible enough, however, that tells us nothing about why those things that can value *must* value. Nor *a fortiori* does it tell us why moral agents must value.

A second *apparent* implication of the conditionality thesis is that life is a sufficient condition of valuation: anything that is alive does indeed value. But reflection should cast some doubt on this implication. For one, it applies only to the proper subset of conditional entities that act deterministically, not to free agents. Given the conditionality thesis, it will be true of any deterministic organism that if it is alive, it is bound to be engaged in valuation. But moral agents are *ex hypothesi* free agents, and I’ve above suggested that as free agents, we are free not to value. In that case, of course, our
being alive cannot be a sufficient condition for valuation—the free choice not to value being an obvious counterexample. So, strictly speaking, life is a necessary but not (quite) sufficient condition for valuation.

A closer look at the issue of sufficient conditionality, however, sheds light on the real point of the conditionality thesis. Let’s put free agents aside for the moment, and restrict our attention to those conditional entities that lack the capacity for metaphysical freedom—for all intents and purposes, to all organisms besides human beings. In that case, it becomes apparent that (though true) it badly understates things to say that “life is a sufficient condition of valuation.” To say that is merely to say that if \( x \) is alive, \( x \) does in fact value. But that formulation misses a crucial modal dimension of the thesis: the point of the thesis is not merely that if \( x \) is alive, \( x \) does value, but that \( x \)’s being alive requires \( x \) to value; in other words, if \( x \) is alive, \( x \) must value in order to sustain its existence and identity.

The implication is that a conditional entity has to set goals to meet its needs, and has to meet its needs to remain in existence, satisfying higher-order needs once the initial needs have been met, and yet higher-order needs once the intermediate needs have been met. Failure to satisfy needs undermines the entity’s existence in proportion to the urgency of the need. Such an entity faces a constant alternative, and the constant necessity of appropriate action in the face of that alternative. “Appropriate action” is action that takes life as the ultimate basis of necessity and selects ends and means accordingly.

The conditional character of life thus confronts an organism as a constant demand for a specific set of actions. In doing so, it constitutes what scholars in Aristotelian
studies call an inclusive end—an all-encompassing end constituted by a set of goals which set is in turn a means to itself.\textsuperscript{12} In the nature of the case, life is both an inclusive and selective end: it is constituted by a set of goals, but it also serves to determine the appropriateness of a given set. Survival-conducive goals can constitute life as an inclusive end, but malconducive goals will not. Notice that since life is an inclusive end, it encompasses the organism’s full lifespan; so the goals that constitute it are the organism’s lifespan interests, the goals that retain their identity across the organism’s lifespan, and in doing so, preserve the organism’s identity as well.

6.3.3 Laying the Foundation (2): The Conditionality of Life Applied to Moral Agents

The account in the preceding section tells us how conditionality applies to agents that \textit{lack} freedom. But what about \textit{moral} agents—agents that paradigmatically \textit{have} it? Life cannot in their case literally be a sufficient condition for valuation; a free agent is (metaphysically) free to ignore conditionality and its implications. He is free, in fact, to flout them. In that case, \textit{nothing} can be a sufficient condition for valuation but the choice to value itself. And as a free choice, that choice cannot be a causal consequence of the conditionality of life. How then is the conditionality of life relevant to it? How does it explain why an agent must value?

The short answer is that it doesn’t. Indeed, it does just the reverse. The conjunction of the conditionality of life and libertarianism tells us that \textit{nothing} explains why a free agent \textit{must} value, if by “must” we’re seeking a categorical requirement. What it explains instead is the basis of what is valuable \textit{if} the agent chooses to value anything. What is true in the deterministic case is true as well in the case of free agents: life is a
process of self-sustaining and self-generated action; it is an inclusive and selective end constituted by a set of goals; it partitions the world into the good, the bad, and the indifferent; it confronts us with a set of needs that demand action if they’re to be satisfied, with each set of needs giving rise to another set once they are satisfied; our existence and identity depend on teleologically appropriate action relative to life so understood; and default on teleologically appropriate action undermines our existence and identity in proportion to the magnitude of the default.

The answer to our question, then, is this: Ultimately, I must engage in valuation because as a living being, I’m a conditional being, and as a conditional being, my existence and identity depend on self-generated and self-sustaining action. Nor does it ever stop depending on it. If life requires continuous action (as *ex hypothesi* it does), then every need I satisfy will bring another one in its wake, at a higher level of complexity. Whatever the variety and complexity of that structure, each element in it gets its *raison d’etre* from its contribution to my life. Default on any element is default on the specifically human mode of living. I can ignore or evade that fact about myself, but I can’t escape it, at least as long as I choose to value and aim to exist. If so, I have to identify what to pursue, and how to pursue it; I have to discover what’s valuable and resolve to track it.

I take this account, though compressed, to answer the version of the “Why be moral?” question that I set out in Chapter 5 and repeated above in this chapter. The question, recall, was: what is the justification, compatible with internalism and metaphysical libertarianism, for why a rational agent must rank alternative courses of action and act on the rankings that he or she sets?
The most succinct answer is that the agent’s existence and identity—and by implication, its needs and interests—depend on doing so. A moral agent is a free agent (libertarianism) capable of limiting its options when it rationally judges that it is right to do so (internalism). Being free, an agent has the option to limit its options or not; having limited them, the agent has the choice to limit them in a specific way or not. The conditionality of life gives the agent a powerful justification for doing both things. A free and rational agent must choose options, and close off others to preserve itself as a free and rational agent. As the conditionality thesis implies, some options preserve the agent’s life, others destroy or impair it. So, self-preservation requires that the agent choose the former and volitionally foreclose the latter. But since conditionality applies to the whole of the agent’s lifespan, a moral agent must do this consistently across the whole of its lifespan: he must rank alternatives to track their value as set by conditionality and act on the rankings so as to preserve his existence and identity. As I’ll argue in the next few sections, the activity of ranking alternatives so understood and acting on the rankings leads us to a virtue-based form of ethical egoism.

6.3.4 Deriving the Superstructure (1): From the Conditionality of Life to a Virtue-Based Act Egoism

As described in the previous sections, conditionality is an attribute of all and only organisms—from the simplest prokaryote all the way up and across the biological taxa. Since humans are organisms, human beings are therefore conditional beings like all the others. Of course, as a distinctive type of organism, human beings also differ from all other organisms in significant ways. The application of the conditionality thesis to the
human case is thus a complicated affair. In one sense, the thesis has a generic application to humans in virtue of the ways in which humans resemble other organisms. In another sense, the thesis has a species-specific application to humans in virtue of the ways in which humans differ from other organisms. Both facts simultaneously have to be borne in mind in applying the thesis to moral contexts and deriving the superstructure of a foundationalist ethics.

As we’ve seen, the conditionality thesis makes three claims about organisms that apply, at a generic level, to the human case. The first is that every organism has an essence; the second is that every organism’s interests are to be individuated in terms of a basic set of lifespan interests; and the third is that the organism’s preservation of its essence is both the means to the realization of its lifespan interests and itself the premier lifespan interest. The question is how these claims apply to the human case.

In non-human animals, valuation is a thoroughly deterministic phenomenon, genetically coded by natural selection, and triggered by features of the environment (including other organisms). Whatever their simplicity or complexity, the lives of non-human organisms proceed more or less automatically—from metabolism, homeostasis, and growth through locomotion, learning, and cooperation. Non-human organisms are in short automatic deterministic value-trackers with life as their ultimate value, and an automatic awareness of, and propensity to act on, their needs. The problem of the conditionality of life is for them solved by genetic coding. They act so as to preserve their lives because they can do no other.

By contrast, for human beings, and by implication for moral agents, valuation is a matter of rational agency engaged in without innate ideas or prescriptions. Whereas
non-human animals act deterministically, rational agents have to exert initiative in order
to act, and resolution in order to keep acting. Whereas non-human animals have relatively
automatic means of discovering their life-requirements, and of pursuing them, human
beings don’t. Lacking innate behaviors or traits of character, we have to cultivate a willed
commitment to discovering and tracking the good. Our distinctive moral task, then, is
freely to employ reason in the service of both ends. The problem posed by the
conditionality of life is for us a matter of explicit rational codification—genetic coding
being insufficient either for motivating action or for directing it in a life-sustaining way.
Rational agency is thus the specifically human essence, its preservation being both
instrumental to and constitutive of a human being’s life.

The application of the conditionality thesis to the human essence gets us from
meta-biology to ethics. An organism’s preservation of its essence, I noted in a previous
section, will itself be its fundamental lifespan interest. And so it is in the human case: an
agent’s preservation of his own rational agency is his fundamental lifespan interest. But
in the nature of the case, this task is colossally complex. For one thing, lacking innate
knowledge, we have to discover that our agency has to be preserved by specific means,
and discover as well how to preserve it. Second, we have to contend with the sheer
duration of a life: a full human life consists of the total expected length of a human
lifespan, some eighty or ninety years; we thus have to project the task of self-preservation
decades into the future. In doing so, we have to take stock of uncertainty, risk, costs, and
path-dependency—and to do so while bearing in mind that we get only one chance at
success. Third, the values we seek across this lifespan cannot successfully be sought as
though they were a series of disaggregated and unrelated sequences; they have to be
conceived, instead, as an integrated whole, each of whose parts contributes to the sum. Fourth, human beings are reflective organisms with a conception of, and need for, self-worth. The lives we lead are ones that we reflect on and evaluate, and the worth we ascribe to our purposes affects what we can pursue and how. Finally, as social beings, we interact with others. We have to devise a life-plan that allows us to derive maximal advantage from the benefits of social interaction compatibly with recognizing the dangers inherent in it.

In short, the task of preserving rational agency requires a life-long *exercise* of rational agency, and so it requires a code of values. A code of values, so understood, is a set of principles that aims to maintain an individual’s existence and identity by maintaining the individual’s adherence to a life-plan that takes stock of the five facts identified in the preceding paragraph. The idea of a “code” in this context captures two things: (a) the systematicity of a set of principles geared to realizing specific purposes, and (b) the traits required to follow the principles in the concrete circumstances of the world. The principles of a code so understood take the form of what might be called fundamental prescriptive generalizations, identifying the values that preserve and unify a full human life. The traits take the form of self-beneficial virtues, geared to achieving the values in question—but above all, geared to preserving and protecting the conditions of rational agency.

Though this is not the place for a full account of virtue and the virtues, we can I think infer a few of them from the preceding account. Obviously, given the centrality of rational agency to the view, *rationality* will have to be one of the virtues. Given the demands of exercising rationality in the concrete circumstances of life, some versions of
honesty and integrity will also be virtues. At a minimum, a virtue of honesty would facilitate a commitment to the reality principle while protecting the agent from self-deception; integrity would facilitate the integration of the agent’s aims while protecting him from moral weakness or backsliding. Given the centrality of self-maintenance to the view, some virtue of productivity seems called for, as does a virtue of proper pride, given the need for self-evaluation and the importance of judgments of self-worth. Finally, social virtues will be needed, including a master virtue—call it justice—to coordinate all of the social virtues and make judgments of the rightfulness of claims based on their worth.

Interestingly, virtue so conceived provides a rationale for the unity of virtue thesis: the virtues on this account are unified by their contribution to the preservation of rational agency, and so, to self-preservation or survival qua human.16

The view just described is in my view best characterized as a virtue-based ethical egoism. Now, ethical egoism may be defined as the view that a moral agent ought to be the ultimate intended beneficiary of every action he takes (and that he ought to act in the first place so as to bring about self-benefit).17 Ethical egoism so construed is really just the first-personal expression of the conditionality thesis. Suppose that I accept the conditionality thesis. In doing so, I make survival qua human my end. Given the nature of survival as described above, the end in turn requires adherence to an essentialist conception of rationality and virtue grounded in self-benefit (understood a certain way). Let’s call a life lived in this way a flourishing life, and the activity of living it, flourishing.18

It is a trivial implication of an ethic of flourishing that I am morally obliged to adopt motivations (acts, traits) consonant with the conditionality thesis, and reject all
motivations (acts, traits) at odds with it. It is obvious, I think, why egoism is consonant with flourishing, while the motives of duty or altruism are at odds with it. If my own good is my ultimate reason for acting, I have no reason to constrain my actions by norms positively subversive of my good, and no reason to sacrifice my good to others (or things) who (or that) make no causal contribution to it. And so it is a relatively trivial implication of flourishing as I conceive it that it is at odds with altruism and deontology. It is perhaps a less trivial implication that flourishing so understood is incompatible with both satisficing and supererogation. A satisficer aims at the second-best option, but the conditionality of life requires the agent to aim always at what is best for him. Supererogation asks us occasionally to go “beyond what duty” demands, but on an egoistic view, duty always demands the same thing: that we bring about the action of the available alternatives that we take to be best for us.19

Though flourishing is a virtue-based egoism, it is—perhaps counterintuitively—an act egoism. An act egoism tells us that every act the agent initiates ought to aim at self-interest. A rule egoism tells us that the agent ought to espouse rules that would generally bring about self-interest, but doesn’t tell us why it is in the agent’s interest to accept the rules. A virtue egoism tells us that the agent ought to acquire traits of character that generally secure his self-interest, but doesn’t tell us why it is in the agent’s interest to acquire those traits of character.20 The view I have defended here is a thoroughgoing act egoism: every act, including the decision to inculcate traits of character, ought to aim at the agent’s interest. It just so happens that the agent’s interest is such as to require the inculcation of virtues, and then to act by expressing them. So while egoistic agents as I understand them are morally required to act virtuously, there is no gap here to be filled
between act, virtue, and self-interest: the expression of virtue is in the agent’s self-interest, as is the act of acquiring it. Hence the somewhat tedious description of the view as a “virtue-based act egoism.”

Finally, egoism so construed bears a complex relationship to consequentialism. There are at (the very) least five distinct and often conflated definitions of “consequentialism” in the literature.\textsuperscript{21} For our purposes, two are relevant. In one sense, consequentialism is the doctrine that right action consists in the maximization of some non-moral good. In another sense, consequentialism is the doctrine that right action consists in bringing about the action from the range of alternatives that will produce the best causal consequences. Egoism is consequentialist in the latter, not the former sense.

It is worth noting, however, that a consequentialism of this latter sort is not a maximizing conception, and so cannot justify doing evil for the sake of a greater good.\textsuperscript{22} The virtues, recall, give us an automatized grasp of the requirements of survival qua human. In doing so, they call for a principled commitment to virtue across the whole of one’s lifespan, a commitment that requires the agent to make principled action an ineliminable part of his moral identity. As Tara Smith aptly puts the point, this commitment to virtue is strong enough to rule out the (apparent) value of “ill-gotten gains” derived from violations of moral principle. Something is “in a person’s interest only if it offers a net benefit to the person’s life. Since a person’s life is not reducible to any isolated element of his condition, we cannot fasten on such elements to draw conclusions” about his interests.\textsuperscript{23} The requirements of virtue, then, are always part of the calculation of “net benefit,” and they override (rather than merely outweigh) vice. So this
momentary departure from principle—this isolated, secret act of vice—no matter how tempting or apparently beneficial, is never in my interest.

6.3.5 Deriving the Superstructure (2): Is Egoism Predatory?

I have so far argued in a very abstract way about egoism’s foundation in the conditionality of life, about its structural features, and very generally about the place of virtue in the theory. I have (by design) said nothing at all about the content of the virtues, which, it seems to me, would require a separate study. My aim here has been simply to sketch the earliest moves that an egoistic ethic must make, leaving it to a subsequent effort to make the later ones. A critic could, however, pre-empt the whole endeavor right from the start by asserting that egoism leads to a reductio: Egoism counsels us to act in our self-interest. But the promotion of our self-interest may well require systematic violations of the rights of others (or generally, of the claims of justice). Since no plausible moral theory can require this, egoism is not a plausible moral theory, even as a candidate for preliminary discussion. Since it isn’t, there is no need to follow the argument any further than we have (if there was a point in following it so far), nor any need in waiting for the sequel. Egoism should be summarily dismissed as a legitimate moral theory.

The cogency of this argument turns on a question nowhere addressed by those who propound it. Egoism does, of course, counsel us to act in our self-interest. But it is hardly obvious that our self-interest does require the systematic violation of the rights of others. In fact, that puts the point far too weakly. As far as I can see, not a single proponent of the preceding objection has explained—or even attempted to explain—why egoism might require a single violation of rights or justice. This is because not a single
such proponent has offered explicit criteria for what counts as an instance of “self-interest” in the first place. The tacit assumption that underwrites the entire literature of criticisms of egoism seems to be that egoistic interests have their source in unconstrained preference-satisfaction, and so involves a limitlessly flexible conception of “interests,” as expressed, say, by a Humean conception of reasons or R.B. Perry’s axiology.\textsuperscript{25}

Abstract as the preceding account of egoism has been, it has the merit of challenging such assumptions at their root. Egoism is not, as its critics have imagined (and asserted), a free-standing ethical commitment that floats free of broader epistemic or axiological commitments. It is, obviously (at any rate, I hope that it is obvious), not an expression of a preference-satisfaction conception of value, nor is it compatible with a limitlessly flexible conception of “interests.” Egoism is instead a consequence of a much broader normative conception—a particular instance of the conditionality thesis, itself a particular instance of a commitment to (a certain brand of) meta-ethical foundationalism. Given this, it gets its content from a non-preference-based conception of flourishing rooted in an essentialist conception of life, needs, interests, and virtues. Objections to egoism that fail to address this broader normative conception simply fail as objections to egoism properly understood. They are not objections to the sort of egoism I have defended here; they either beg the question or are entirely irrelevant to it.

I cannot, in what follows, offer a full-scale account of the nature of self-interest or the content of an egoistic theory of justice.\textsuperscript{26} What I’ll do instead is—given the foregoing account—to lay out the adequacy-conditions for what might count as a genuine counterexample to egoism, explaining why none of the counterexamples described in the literature meets the relevant conditions. Each of them is in fact a textbook-like instance of
some informal fallacy—typically, of begging the question, *ignoratio elenchi*, and/or false alternative.

A first issue harks back to the opposition between optimistic and pessimistic conceptions of personhood and agency discussed in Chapter 5 (5.4.3.1). I argued there that on an optimistic view, there are no conflicts of interest between egoists acting egoistically: it is in my interest to achieve an intrapersonal concord of interests, and (insofar as possible) in my interest to interact with agents with the same aim (insofar as we share aims). Under such conditions, the optimistic conception holds that egoistic social interaction is a “positive-sum game” that holds out the possibility of a harmony of interests in which the virtue of each agent becomes an asset for that of every other virtuous agent. The harmony is underwritten by the fact that each virtuous agent sees his benefit in the virtues rather than the vices (or involuntary weaknesses) of others. A virtuous egoist benefits neither by sacrificing his interests to others, nor by preying on their vices or vulnerabilities, but by interacting with others in order to gain from their virtues and strengths, and by implication, from their self-interest. The rational pursuit of self-interest, then, is never the cause of conflicts of interest, because virtue constitutes our self-interest, and virtue is never the cause of conflict. Conflicts arise not through the pursuit of objective interests, but through violations of the egoistic principles that secure our interests.

The optimistic conception conflicts, of course, with the pessimistic conception we find in the literature. Egoism, Baier tells us, is “among the main causes of our social problems.” But Baier evinces little appreciation for how large a generalization this is, for what kind of generalization it is, or for what would confirm or disconfirm it. In
particular, he provides no evidence for the generalization, apart from adducing discrete examples of conflict situations. Isolated examples don’t prove that something is “among the main causes” of something else; at best, they serve to confirm the initial impression that it is. So Baier’s argument begs the question. In any case, Baier misses the fact that the egoist claims that egoistic interests don’t conflict. That claim is disputable, but Baier does not dispute it. He simply assumes that egoistic interests do conflict, and counts this as a counterexample to egoism. But this claim begs the question against what egoists really do claim about self-interest, and constitutes an ignoratio of what can be established by way of claims about conflict. A successful counterexample to egoism would have to address the optimistic conception presupposed by egoism.

A second issue involves the fact that egoistic interests are lifespan interests, so that all genuine egoistic interests are interests that span a full lifetime, or are individuated in terms of their causal contribution to such interests. This fact precludes the use of counterexamples like this one, offered by James Rachels:

The ethical egoist would say at this point, “Of course it is possible for people to act altruistically, and perhaps many people do act that way—but there is no reason why they should do so. A person is under no obligation to do anything except what is in his own interests.” This is really quite a radical doctrine. Suppose I have an urge to set fire to some public building (say, a department store) just for the fascination of watching the spectacular blaze: According to this view, the fact that several people might be burned to death provides no reason whatever why I should not do it. After all, this only concerns their welfare, not my own, and according to the ethical egoist the only person I need think of is myself.\textsuperscript{28}

Rachels unaccountably jumps here from “S ought to act in his interest,” to “S ought to act on his urges,” to “S ought to burn down a building and kill everyone inside,” concluding that since the fire concerns the occupants’ welfare, it doesn’t concern S’s. But Rachels
has given us no reason for thinking that any particular thing is in an agent’s interest, much less that any particular urge is in the agent’s interest, much less that all urges are, much less that this urge is. As we’ve seen, since egoism is not a preference-satisfaction theory, an appeal to “urges” is an obvious *ignoratio elenchi*. So too is the claim that the occupants’ welfare is unrelated to S’s: the example doesn’t tell us whether the two things are connected or not.

More to the point, however, Rachels’s counterexample turns on a complete disregard for the concept of a lifespan interest. A rational egoist as I’ve described him takes his life as a single unit of long-term deliberation, individuating particular interests by their contribution to life as an inclusive end. He does not think of his interests in terms of discrete and momentary sequences unrelated to the rest of his life. So apart from the fact that egoism is not a preference-satisfaction theory, Rachels’s appeal to an “urge” for arson gets things wrong yet another way: there is no legitimate inference from *any* range-of-the-moment phenomenon (whether cognitive or appetitive) to what is in the agent’s objective interests. The inferences, rather, go the other way around: we need to know what is in the agent’s *lifespan* interests to know, in a given situation, what is in his interests here and now. A successful counterexample, then, would have to explain why a putatively immoral trait or act was either a lifespan interest or contributed to one. Rachels makes no attempt to do this, and neither does anyone else.

A third issue involves the place of virtue in discussions of egoism. Without specifying the content of virtue, I have argued that egoism implies that virtue is in our interest. There is, however, little or no discussion in mainstream counterexamples of the place of virtue in an egoist’s lifeplan. When the egoist’s values are not straightforwardly
depicted as psychopathological, discussions of egoism almost always concern the unjust seizure of goods (or exploitation of persons for this purpose). Such claims miss the fact that on an egoistic view, virtue is a lifespan interest, so that goods are only valuable when gotten by methods expressive of virtue. Thus James Sterba offers a fairly typical counterexample to egoism based loosely on the counterexample in Book II of Plato’s *Republic:*

Consider the case of Gary Gyges, an otherwise normal human being, who for reasons of personal gain has just embezzled $300,000 while working at People’s National Bank and is in the process of escaping to the South Sea Islands, where he will have the good fortune to live a pleasant life, protected by the local authorities and untroubled by any qualms of conscience. If we assume that in the society from which Gyges is fleeing moral standards are generally observed, Gyges’ behavior would be obviously immoral. But, Sterba suggests, it seems an unexceptional expression of egoism. This example, like the two preceding it, simply assumes without argument that embezzlement is in Gyges’ interest, and assumes that there are no countervailing considerations requiring discussion in this regard. If Gyges wants money, the example implies, money is in his interest, period. It also makes a rather large raft of assumptions about the sheer logistics of the operation.

But put those aside. The real difficulty with Sterba’s example is that it omits any discussion of virtue. Now, nothing I have so far said about egoism tells us that theft is contrary to one’s interest: for all I’ve said, it might well be in one’s self-interest; that matter can only be resolved once we have in hand a theory of interests. But we can infer from what I’ve said that theft would be in a person’s self-interest only if it were expressive of virtue. Since virtue is in one’s interest, and virtue is a lifespan interest, if
theft is in Gyges’ interest, it must express some virtue that is in his interest—and, indeed, some virtue that expresses the unity of virtue. Those are, as I’ve suggested, structural features of the sort of egoism I’ve defended here.

But Sterba neglects to tell us which virtue or set of virtues is expressed by theft. In neglecting to tell us this, he helps himself to the stipulation that Gyges is “an otherwise normal human being,” who is content with his life in the South Sea Islands, and “untroubled by any qualms of conscience.” Sterba has thus simply assumed without argument—indeed by stipulation—not just that theft is compatible with psychological well-being but that a life expressing a virtue requiring theft is.

The dilemma he faces is this. Either theft expresses an egoistic conception of virtue or not. If not, theft is not in Gyges’ interest, and the thought-experiment immediately falls apart. But if theft expresses egoistic virtue, Sterba’s thought-experiment will only work if tells us (a) which virtue it expresses, and (b) how its expression is consistent with the other virtues. Of what self-beneficial lifespan trait is theft an essential expression? How does that trait cohere with all of the other of an agent’s lifespan traits? Sterba has to answer those questions before asserting that Gyges is “otherwise normal,” and “untroubled by qualms of conscience.” He would only be that way if theft were an expression of human virtue. But in the nature of the case, a paragraph-long thought-experiment does not explain how that is possible, much less why it is plausible. In the absence of such an account, thought-experimental counterexamples against egoism have all the merits of theft over honest toil: they assume precisely what needs to be demonstrated, but demonstrate close to nothing.30
6.4 Empiricism and Egoism: The Structure of a Comprehensive Normative Foundationalism

In this last section, I want at last to explain how the two forms of foundationalism, epistemic and meta-ethical, come together to constitute a comprehensive normative foundationalism: i.e., a single foundationalist theory comprising two distinct but related sub-theories. There are two different tasks here. In this chapter, I’ve defended my own favored conception of foundationalism: Aristotelian empiricism in epistemology, ethical egoism in meta-ethics. We need to see how these two commitments interrelate to form a single coherent theory. But there is a broader point to be made here about foundationalism as such, in abstraction from my favored conception of it. There are, as we’ve seen, two different ways of approaching the foundations of ethics—epistemic and meta-ethical. But it’s obvious that these are not two utterly distinct and completely unrelated approaches; they are partially distinct but obviously related to each other. They are—as the example of Aristotelian empiricism and ethical egoism should make clear—combinable approaches. We thus need to identify the structure common to any comprehensive normative foundationalism regardless of its particular form. How, generally, do epistemic and meta-ethical foundationalisms interact in an inquiry into the foundations of ethics?

Begin with the case defended in this chapter. I have argued that one approach to the foundations of ethics proceeds via a commitment to epistemic foundationalism. In this case, we accept epistemic foundationalism as a general thesis about epistemic justification, and apply it to the case of moral beliefs. At this point, we face several
alternatives, one of which is Aristotelian empiricism. A second approach to the foundations of ethics proceeds via a commitment to meta-ethical foundationalism. In this case, we ask some version of a “Why be moral?” question, answer it, and derive a superstructure of norms from the foundation. At this point, as well, we face several alternatives bearing both on the form of the question we ask as well as the answer. One answer here is ethical egoism.

We could, I suppose, treat Aristotelian empiricism and ethical egoism as though they were two utterly unrelated commitments—two unrelated answers to two unrelated issues. But that is implausible. The two commitments may differ, but being about the same subject-matter—the justificatory basis of ethics—they cannot be entirely unrelated. Suppose now that our epistemology, though foundationalist, involves a coherence proviso: we should in general aim at as much coherence as is compatible with foundationalist strictures. In that case, our (foundationalist) epistemology gives us an incentive for actively seeking out the connection between these ostensibly unrelated theoretical commitments—and to exploit any connection we find so as to achieve coherence.

One way of doing this would be to ask whether Aristotelian empiricism was merely consistent with ethical egoism. Obviously, if the two commitments are inconsistent, a coherence proviso will require us to drop one. But if the two commitments are consistent and related to one another, as they seem to be, a coherence proviso will require us to adopt what we may call an integrated conjunction of the two commitments. Since the two commitments are consistent, we can in principle believe their conjunction. But believing the conjunction of two consistent propositions does nothing to advance the
coherence of the believer’s belief-set: if \( p \) and \( q \) are consistent, \( S \)’s believing \((p \& q)\) has no net effect on the coherence of \( S \)’s belief-set. To increase coherence, \( S \) needs to inquire into whether \( p \) and \( q \) are mutually supporting, and if so, to espouse them as mutually supporting. Doing so will require identifying the implications of both \( p \) and \( q \), and showing connections not only between \( p \) and \( q \) but also between the various implications.

So it is Aristotelian empiricism and ethical egoism. Aristotelian empiricism is a form of epistemic foundationalism. If this foundationalism includes a coherence proviso, it will require a foundationalist not just to conjoin but to integrate commitments to empiricism and egoism.

The total structure may be described as follows. We begin with a commitment to epistemic foundationalism. That leads \((ex \ hypothesi)\) to Aristotelian empiricism, which is applied to the case of moral belief. Aristotelian empiricism gives us an epistemic foundation of sensory perceptions and tells us that moral claims rest in a superstructure atop these perceptions. We get from the foundation to the superstructure by abstraction. But abstraction presupposes an account of similarity: to form moral concepts from the perceptual level, we need to know what it is that those concepts have in common.

That pushes us to the meta-ethical approach to the foundations of ethics. In formulating a “Why be moral?” question, and answering it, we come to an account of precisely what it is that unifies moral concepts: the answer to that question is supplied by the answer to the “Why be moral?” question. In the case of ethical egoism, moral claims and moral concepts are unified by their contribution to self-interest, which itself gets its content from the conditionality of life. So, in abstracting concepts from the perceptual
level, we rely on our meta-ethical foundationalism to know where to look and what to abstract.

As an epistemic commitment, Aristotelian empiricism tells us that moral claims are justified by a foundation of sensory perceptions. If the meta-ethical foundation directs us to look at concepts of self-interest and virtue, the combination of the epistemic and meta-ethical commitments tells us that the foundation of ethics is to be sought by deriving claims about self-interest and virtue from the evidence of the senses.

In reflecting on this structure, we can see that epistemic and meta-ethical foundationalism are obviously not merely homonyms. They are complexly related to each other. Nor are they merely analogous. They are too intimately related for that. Nor is meta-ethical reductionism involved: the meta-ethics does not drive the epistemology. We could say, however, that the meta-ethical account is reducible to the epistemic. It’s the epistemology that drives the account as a whole: the account as a whole has to square with epistemic foundationalist strictures. And so there is a case to be made that the preceding structure instantiates epistemic reductionism. But the meta-ethical part of the account has a certain autonomy from the epistemic while being analogous to it: the “Why be moral?” question can be answered independently of a commitment to epistemic foundationalism, and has to be answered so as to get the very task of abstraction off the ground. The result, then, is a complex version of the structure that I called an “epistemic reductionism and analogy hybrid” in Chapter 1 of this study. The meta-ethical approach to the foundations of ethics is reducible to the epistemic approach in the sense that the meta-ethical account must—like all our beliefs—be justifiable by foundationalist
strictures. But it is in another sense autonomous of, hence merely analogous to, epistemic foundationalism as a theory.

What is true of my own favored account of foundationalism will, mutatis mutandis, be true of any comprehensive normative foundationalism. In each case, we begin with a commitment to epistemic foundationalism. This commitment tells us where justified moral belief ends up in the overall structure of justified belief. It also tells us that our beliefs (including our theoretical beliefs) must, consistently with foundationalist strictures, be maximally coherent. And so it gives us an incentive to integrate our moral epistemology with our meta-ethics. The meta-ethics, in turn, tells us how to answer the “Why be moral?” question, and in so doing, tells us what moral propositions have in common: they promote our self-interest, they are universalizable, they promote overall utility, they comply with divine commands, they fulfill our contractual promises, etc. Once we answer the “Why be moral?” question, we see what moral claims have in common and derive a superstructure of norms atop the foundation. The meta-ethical account, of course, has to be squared with the epistemic one. In doing so, it has to show why an account that answers the why be moral question is epistemically justified, and so truth-conducive. The result, in each case, is (ideally) a comprehensive normative foundationalism that tells us why we should be moral while confirming confidence in our justified moral beliefs, casting doubt on the unjustified ones, and allowing us to track the truth about morality as a whole.\textsuperscript{31}
6.5 Notes


Though Rand’s argument has not gotten much attention in academic philosophy, one exception is Robert Nozick’s (much-vaunted but risibly deficient) “On the Randian Argument,” reprinted in Robert Nozick, Socratic Puzzles (Cambridge: Harvard University Press, 1997), pp. 249-64. Nozick claims, disingenuously, that his essay “examines her moral arguments, which academic philosophers previously had neglected, but since it treated with the standard critical tools of the trade, it brought me vituperation from some of Rand’s followers” (p. 6). In fact, Nozick’s essay cannot be said to examine either Rand’s arguments or much of anything at all: the “standard tools of the trade” to which he alludes include sixty-four rhetorical questions—along with a dismissive series of misattributions, irrelevant thought-experiments, and obvious fallacies. Among the responses that it “brought…from some of Rand’s followers” is one outright refutation that Nozick assiduously ignores in his subsequent writing (Douglas Den Uyl and Douglas Rasmussen, “Nozick on the Randian Argument,” reprinted in Reading Nozick: Essays on Anarchy, State and Utopia, ed. Jeffrey Paul [Totowa, NJ: Rowman and Littlefield, 1981], pp. 232-69). Den Uyl and Rasmussen convincingly establish within the first eight pages of their essay that Nozick has simply not mastered the argument he claims to have refuted.

Nozick blithely tells us that “I have not responded to the sizable literature on Anarchy, State, and Utopia, or followed it closely…I have not responded to the literature on my other writing either…I could say that any intellectual contribution I could make was likely to be greater in broaching new ideas on new subjects than in elaborating or defending my previous ideas…” (Nozick, Socratic Puzzles, p. 2). Apart from the utter arbitrariness of the latter claim, justified on the basis of nothing but whim, it has no application to “On the Randian Argument,” which purports to be a critique of Rand’s ideas, not an exploration of Nozick’s.


I have omitted a discussion in this chapter of Michael DePaul, Balance and Refinement: Beyond Coherence Methods of Moral Inquiry (London: Routledge, 1993). The omission, though deliberate, is not intended to minimize the book’s relevance to the topics discussed here, or its importance. The rationale for the omission is this: From my perspective, DePaul conceives from the outset what I take to be the most damaging objection to either reflective equilibrium or a method of balance of refinement—that neither reliably track moral truth (pp. 55-56). (He occasionally lapses into writing as though the issue were “guaranteeing” truth, but the point is that neither method is reliably or strongly truth-conducive.) Despite the ingenuity and interest of the rest of his theory, I take the preceding to constitute a reductio. As I read Balance and Refinement, the crux of DePaul’s theory comes in chapter 2’s discussion of the goal of moral inquiry, and of rationality in chapter 3—topics (and disagreements) that would take me too far afield to discuss in the present context.


For the relevant (i.e., relevantly essentialist) sense of identity, see Jonathan Jacobs, “Teleological Form and Explanation,” in Current Issues in Teleology, ed. Nicholas Rescher (Lanham, MD: University Press of America, 1986).

I’m using the phrase “traits of character” here in the broadest sense possible—neutrally as between deterministic and libertarian ways of acquiring traits of character, and also neutrally as between voluntary and involuntary accounts ways of acquiring them. Hence in the sense I intend, some organisms acquire traits of character freely and responsibly in the libertarian sense; some acquire them voluntarily (in the compatibilist sense) but not freely (in the libertarian); and some may well acquire them involuntarily and deterministically. Thus persons have traits of character in the intended sense, but so do gorillas, dogs, cats, gerbils, parrots, camels, and manatees. I also use the term “trait” here in a broader and more general sense than that used in biology. Biologists define traits as “a particular phenotypic character as opposed to character mode” (M. Abercrombie, M. Hickman, M. L. Johnson, and M. Thain, Penguin Dictionary of Biology 8th edition with Amendments [New York: Penguin, 1992], p. 562). As I am using the term, “traits” subsume both particular characters as well as character modes, and subsume both physical characters as well as psycho-physical ones. I am also abstracting from the difference between traits determined by phenotype and genotype, and traits cultivated by agent-causal volition.

Abercrombie et al., The Penguin Dictionary of Biology, p. 402.


This claim may seem vulnerable to obvious counterexamples. (1) For one thing, most organisms engage in reproductive behavior, which is costly relative to survival and yet engaged in anyway. (2) Second (and relatedly), some animals engage in suicidal or altruistic behavior (e.g., male praying mantises, worker bees), which would seem to be a counterexample to the claim that organisms act deterministically to preserve survival. (3) Third, it might be argued that surely natural selection operates in order to preserve genes, not individual organisms. And if natural selection gives us our theory of goals—Darwinian theory being our best theoretical account of natural goals—a Darwinian theory of goals based on natural selection will override a non-Darwinian one not so based, so
that gene-preservation will override self-preservation. For all three objections (though not phrased there explicitly as objections to the present view), see Richard Dawkins, *The Selfish Gene*, 30th anniversary ed. (Oxford: Oxford University Press, 2006).

Common as these claims are, however, none of them to my mind succeeds as a counterexample to the view defended in the text. The first two objections are (as far I can see) directly and conclusively answered in Harry Binswanger’s underrated and underread book, *The Biological Basis of Teleological Concepts*, pp. 153-59; the third objection is answered indirectly on pp. 159-63.

With respect to the first two objections, Binswanger points out (what would appear to be the obvious fact) that reproduction confers a benefit on an organism by bringing the organism into existence. Since (given genetic determinism) “there is no way for a given organism to obtain the benefits of past acts of reproduction without having to repeat those acts upon their own maturity” (p. 156), the costs of reproduction are simply a (sunk) cost of (the immense benefit of) having come into existence. Given this, the costs of reproduction merely constrain the way in which the organism preserves its life (as do the costs of metabolism or excretion); their existence doesn’t rebut the claim that self-preservation is an organism’s ultimate end. That answers objection (1), and all instances of objection (2) where the “altruistic” behavior contributes to (and arises from) reproductive activity.

On pp. 159-63, Binswanger offers a similar argument against the remaining cases of (2) not covered under the preceding objection, and though I lack the space to argue for the claim here, I think that the preceding arguments can, *mutatis mutandis*, be deployed against objection (3).

Objection (3) is in any case complicated by Dawkins’s exposition. On p. viii of the 30th Anniversary Edition of *The Selfish Gene*, he tells us that the topic of his book is “what kind of entity is it that survives, or does not survive, as a consequence of natural selection.” On the same page, he answers that “individual organisms” behave “altruistically ‘for the good of the genes’.” These two clauses—one naming the book’s question, the other its answer—illustrate a confusion that pervades the book. First, something’s being a consequence of natural selection is not the same as its being a goal, in which case, Dawkins’s thesis doesn’t count as a counterexample to the view defended in the text. But in any case, I would argue that genes are not conditional entities (and genotypes are not entities at all), so that they do not count as genuine beneficiaries of anything. In that case, it does not make sense to say that an organism acts “for the good of genes,” anymore than a hydrogen ion acts “for the good of” a water molecule. Genetic replication could simply be a consequence of survival-aiming behavior, not its goal, in which case the counter-example misfires.

These are, however, complex topics in the philosophy of biology beyond the scope of this study, and so deserve a separate treatment. Thanks to James Lennox, Michael Young, Jason Raibley, and Carrie-Ann Biondi for helpful discussion on this topic.

14 I have assumed here without argument that innatism is false, at least in any sense that would allow for any innate knowledge relevant to self-preservation or anything
that might otherwise qualify as innate moral knowledge. Rejection of nativism would take more argument than I have space for here, but a few brief remarks are in order.

A first difficulty with innatism concerns its proper formulation: what exactly are advocates of innate ideas or innate knowledge saying that distinguishes them from empiricists? As G. A. J Rogers notes with respect to Noam Chomsky’s conception of the thesis, “But for [Chomsky’s] purposes innate ideas must be taken in a strongly dispositional sense—so strong that it is far from clear that Chomsky’s claims are as in conflict with empiricists’ accounts as some (including Chomsky) have supposed” (G. A. J. Rogers, “Innate Ideas,” in A Companion to Epistemology, ed. Jonathan Dancy and Ernest Sosa [Malden, MA: Blackwell Publishers, 1992], p. 217).

A second difficulty concerns the move from a generic assertion of the existence of innate ideas to knowledge or ideas relevant to action-guiding prescriptions. The best hope for such a view seems to come from sociobiology or “evolutionary psychology,” but the evidence offered for it seems to me underdetermined and implausible.

A third difficulty is that even if innate ideas or knowledge existed, they wouldn’t vitiate the claim made in the text: such innate knowledge as we (ex hypothesi) have would not operate in anything like the way non-human animal learning does. In particular, it would do nothing to undermine the claim that we have to discover the knowledge required for survival and take the initiative to systematize and act on it.

I take my rejection of innate ideas to be incompatible with at least one substantive view in moral epistemology: Rawls’s Chomsky-inspired account of “the sense of justice” (John Rawls, A Theory of Justice, rev. ed. [Cambridge: Harvard University Press, 1999], pp. 41-42). See also his earlier Rousseau-influenced account, “The Sense of Justice,” in John Rawls, Collected Papers, ed. Samuel Freeman (Cambridge: Harvard University Press, 1999), p. 96. As Rawls puts it, “In Emile, Rousseau asserts that the sense of justice is no mere moral conception formed by the understanding alone, but a true sentiment of the heart enlightened by reason, the natural outcome of our primitive affection” (my italics). The thesis I defend in the text is incompatible with the thought expressed by the italicized phrase.

15 The phrase is Tara Smith; see Smith, Viable Values, p. 165.


17 I owe this definition to Allan Gotthelf.

David O’Connor has suggested to me that a first-person avowal of the conditionality thesis need not lead by itself to egoism. The conditionality thesis asserts that each of us is a conditional being whose existence and identity depend on our own goal-directed action. Egoism asserts that each of us ought to be ultimate intended beneficiaries of our own action. The argument in the text supposes that preservation of an agent’s existence and identity is beneficial for the agent, and that conditionality requires the agent, in every act, to aim at preserving this existence and identity, thereby bringing about his or her own benefit.

The claim is susceptible of at least two closely related objections. (1) What if there were goods required by the conditional character of life which were such that aiming at them by intending one’s ultimate benefit was incompatible with attaining them? Thus it could be that the conditionality thesis requires us to engage in intimate other-regarding relationships, but it could also be that the psychological requirements of other-regarding relationships were such as to forbid aiming at another’s good for the sake of realizing one’s own. (2) Similarly, what if the goods required by social and political life were common goods of a sort that precluded our making distinctions between the good of one agent as distinct from that of another? Might it not be the case that the good of a political unit was good for that unit, but not in a way that allowed any given agent to discern a causal relationship between his or her individual actions and that good?

A more detailed defense of egoism would, I think, have to deal at length with these objections. For the present context, I would say in response to (1) that I reject the assumption that intimate relationships are subverted by the simultaneous aiming at one’s own good and that of the intimate. After all, if intimate relationships are good for both agents, it is unclear why self-conscious recognition of this fact should undermine the quality of the relationship. Agents might, on various occasions, be dissuaded from an obsessive fixation on their own good, but it’s hard to see why agents should be threatened by recognition of what is after all a fact, even by the terms of the objection—that intimate relationships are mutually beneficial.

Objection (2) invites a similar response. Either the common good benefits every member of a political unit, or sacrifices some. A common good of the latter sort is incompatible with the egoism I defend in the text. If it benefits every member, however, we should be able to identify how it does so, and in doing so, discern the causal relationship between individual actions and mutual benefit. I admit that this might prove difficult, but the more difficult it proved, the more difficult it would be to assert, with confidence, that the goods realized through political life were indeed common goods.

I know of no published discussion of the relation between egoism and supererogation. I have merely indicated in a broad way what I take to be the incompatibilities between egoism and the four views mentioned in the text; obviously, the topic bears further discussion than I can give it here.

The egoistic rejection of altruism (self-sacrifice) disposes of James Sterba’s objection to egoism in James Sterba, Justice for Here and Now (Cambridge: Cambridge University Press, 1998). Sterba asks how an Aristotelian flourishing-based ethic would “explain our experiences of self-sacrifice” (p. 35), suggesting that the experience would either have to be explained as a conflict between selves or between aspects of a single self. I agree that that is how self-sacrifice would be explained. Unfortunately for Sterba’s purposes, on the view I’ve defended here, that explanation would also explain why altruism is immoral.


21 On this point, see my review of Stephen Darwall’s Consequentialism, Teaching Philosophy, vol. 28, no. 3 (September 2005), pp. 282-83.

22 John Cooper makes a similar point regarding Rawls’s definition of “teleological ethics” in Cooper, Reason and Human Good in Aristotle, pp. 87-88. I dissent, however, from Cooper’s criticism of Ross (p. 87 n. 113). Cooper quotes Ross as saying that “Aristotle’s ethics is definitely teleological; morality consists in doing certain actions not because we see them to be right in themselves but because we see them to be such as will bring us nearer to the ‘good for man’.” (See W. D. Ross, Aristotle [New York: Meridian Books, 1959], p. 184.) Though Aristotle’s ethics is not teleological on, say, Rawls’s conception of that term, I see no reason, even on Cooper’s grounds, to take exception with Ross’s formulation.

23 Smith, Viable Values, p. 168. It is, I hope, an obvious implication of this claim and the foregoing account that egoism rules out a desire-based or preference-satisfaction conception of value or reasons for action. Desires or preferences are, precisely, “an isolated element of a person’s condition”: a person is neither identical to a desire nor identical to the sum of his desires, past, present, and future. Since desires do not necessarily track survival-conducivity and are not necessarily expressive of virtue, on an egoistic view, ‘S desires x’ tells us absolutely nothing whatsoever about what is of value to S—not even what is of pro tanto value to S. Conditionality as I have described it is an overriding source of value. It and only it gives rise to the need to act one way or the other. In doing so, it thereby gives a point to desires and preferences. So desires and preferences expressive of genuine needs and interests do of course provide reason for action. But desires and preferences as such do not. I take this issue up again in Chapter 7’s discussion below of Christine Korsgaard’s account of the sources of normativity. But the (essentially Humean) assumption that desires provide pro tanto reason for action is of course widespread in contemporary analytic philosophy.
I thank Michael Young for many constructive disagreements on this topic, and Carrie-Ann Biondi for many useful agreements.

24 I mean “nowhere addressed” in a literal way. In a fairly thorough canvassing of discussions of egoism in the Anglo-American philosophical literature, from 1900 to the present—invoking scores of texts—I have not found a single author who has addressed the issue in even a rudimentary way. The one approximation to an exception I encountered is Alasdair MacIntyre’s entry “Egoism and Altruism,” in The Encyclopedia of Philosophy, vol. (New York: MacMillan, 1967), p. 466, which makes a point that rebuts the objection discussed in the text. As MacIntyre puts it, in the absence of a criterion for what counts as an interest, the terms “egoism” and “altruism” are vacuous. A fortiori, in the absence of such a criterion, the inference from “egoism” to “predation” is an obvious non-sequitur. (I should add that MacIntyre’s treatment of egoism diverges substantially from mine on all points but the preceding one.)


26 Indeed, I’m inclined to say that the abstractness of the argument I have so far is not a liability for egoism but for its critics. I have not offered an account of the prescriptive content of an egoistic ethic precisely because I have needed to spend so much space merely to lay out the foundation of the view. But because the theory’s prescriptions rest on that foundation, critics’ nearly wholesale failure to discuss the foundation is a failure in their criticisms, not a failure of the theory.


28 James Rachels, “Egoism and Moral Skepticism,” reprinted in Vice and Virtue in Everyday Life: Introductory Readings, ed. Christina Hoff Sommers and Fred Sommers (San Diego, CA: Harcourt Brace Jovanovich, 1989), p. 486. In a footnote I’ve omitted, Rachels remarks of the quoted view: “I take this to be the view of Ayn Rand, insofar as I understand her confusing doctrine.” It is quite obvious that Rachels did not understand Rand’s doctrine to even the slightest degree; as we’ve seen in Nozick’s case (note 1 above), failure to understand a doctrine is not an obstacle to criticizing it when the doctrine is Rand’s. Rachels offers no reason to persuade the reader that Rand’s doctrine is “confusing”: indeed, he doesn’t offer the reader a single quotation from a single one of her works.
In a subsequent work, Rachels once again takes on Rand’s argument (James Rachels, “Ethical Egoism,” in *The Elements of Moral Philosophy*, 3rd ed. [Boston: McGraw Hill College, 1999]). Dismissing her (falsely) as “a writer little heeded by professional philosophers” (p. 86), he proceeds immediately to her criticisms of altruism. Thus the larger argument that Rachels once described as “confusing” is now dismissed without acknowledgment or discussion of any kind with the cavalier assertion that “Rand also suggests that there is a metaphysical basis for egoistic ethics” (p. 87). Unfortunately, his explication of this “metaphysical basis” a few paragraphs later does not even mention the basic elements of her argument (pp. 87-88). Rachels then borrows an argument of Baier’s to suggest that it is in an egoist’s interests to kill a rival for a sought-for position (pp. 92-93). Unfortunately, this argument turns on the undefended assumption (in premise [2]) that murder is in an egoist’s self-interest—a claim that follows from the sheer stipulation that the egoist in the Baier/Rachels example has but one interest: to get the sought-for position. Rachels ends his case against egoism by likening it to racism on the assumption that any moral doctrine that assigns greater importance to \( x \) than to \( y \) (where \( x \) and \( y \) are persons) is “unacceptably arbitrary” (pp. 94-95). Apart from the fact that this latter principle is implausibly incompatible with partiality of any kind (e.g., to one’s loved ones), it misses the fact that the conditionality thesis gives us a very good reason for treating our own lives as having more importance than others: as conditional entities, our lives depend for their existence on our own efforts, not those of others; the conditionality of life gives us an obligation to act for our own good, and to treat it as more important (to us) than the good of others (is to us). Rachels’s comparison of egoism to racism misses the fact that racism arguably presupposes both determinism and some version of the doctrine of innate ideas, both of which are incompatible with the defense of egoism I’ve offered, and not coincidentally, explicitly incompatible with Rand’s theory, which Rachels claims to be discussing (cf. Rand, “Racism,” in Rand, *The Virtue of Selfishness*, ch. 17).

29 James Sterba, *The Demands of Justice* (Notre Dame: University of Notre Dame Press, 1980), p. 1. In a later work (Sterba, *Justice for Here and Now*, ch 2), Sterba argues against egoism by suggesting that “high ranking altruistic reasons” ought to take precedence over egoistic ones. But his argument begs the question many times over. (1) For one thing, though he speaks of “rankings,” he offers no actual standard by which to rank either egoistic or altruistic reasons, much less the conflict between them. (2) He admits that his resolution in terms of altruistic reasons “begs the question,” claiming that it begs the question to a lesser degree than does egoism, but this seems to me little more than a confession of failure—and one not reciprocated by the egoist. (3) In any case, there is no coherent account of the notion of committing a fallacy to a lesser degree, and Sterba offers none. (4) Finally, Sterba defines altruism in such a way as potentially to overlap with egoism: altruism, he tells us, can “require little or no sacrifice of self-interest” (my emphasis). But if altruism requires no sacrifice of self-interest, then egoistic reasons can be altruistic and vice versa, in which case the opposition between them is a puzzle, and Sterba’s account of egoism bears no significant relation to the form of egoism I have defended in the text.
One partial exception to this general rule is Bernard Boxill, “How Injustice Pays,” *Philosophy and Public Affairs*, vol. 9, no. 4 (Summer 1980), pp. 359-71. Sophisticated as it is, however, the essay targets Plato’s views, not those discussed in the text, and it thus does not make contact with the claims defended in the text. The injustices Boxill has in mind arise when a group of egoists bands together to exploit or violate a “well-selected group of outsiders.” I assume that the well-selected group of outsiders is chosen by means of identifying principles that violate the principles of justice (or rationality or virtue generally). In that case, however, Boxill ignores the possibility that violating such a principle (and identifying with a group that does) subverts one’s self-interest precisely because adherence to the principle is in one’s self-interest. This will not, of course, convince a proponent of Boxill’s argument insofar as he insists that adherence to such a principle is not in one’s self-interest. But the very form of this disagreement suggests that it cannot fruitfully be carried out on either side in abstraction of a not-merely-antiquarian egoistic account of justice. Since justice is beyond the scope of this study, however, I leave the matter there.

I’ve developed the material in this chapter further in “The Foundations of Ethics: Objectivism and Analytic Philosophy,” in a paper delivered to the Ayn Rand Society, APA Eastern Division Meeting (December 29, 2007), with commentary by Paul Bloomfield.
INTERLUDE:

CHRISTINE KORSGAARD'S THE SOURCES OF NORMATIVITY:

AN EXTENDED STUDY

Christine Korsgaard's *The Sources of Normativity* (hereafter, *Sources*) is an important recent exploration, from a neo-Kantian perspective, of epistemic and meta-ethical issues at the foundations of ethics.\(^1\) Originally delivered as the 1992 Tanner Lectures at Cambridge University, the book has by now been widely discussed in the literature, and its status firmly and deservedly established as one of the standard-setting works of what we might call post-Rawlsian neo-Kantianism.\(^2\)

My discussion here has two interconnected aims. There is first just the intrinsic interest of examining a well-crafted discussion of an important issue to assess the merits of the argument. But there is additionally the fact that Korsgaard's view provides the perfect test-case for the claims I've made in the first part of this study. In previous chapters, I distinguished epistemic and meta-ethical approaches to the foundations of ethics, defended the legitimacy of one motivation for the latter, and suggested the possibility of integrating epistemic and meta-ethical approaches in one comprehensive theory. While Korsgaard discusses many of the same issues as I do, she comes at those issues from a very different perspective. She is, in the terminology introduced previously.
in this study, an anti-foundationalist about justified moral belief, and a weak neo-Kantian foundationalist about meta-ethics. This view is an important rival to the view I myself have defended here, namely, Aristotelian empiricism about justified moral belief and a strong ethical egoist foundationalism about meta-ethics. How well do Korsgaard’s commitments fare as an approach to the foundations of ethics in comparison to mine?

In the next two chapters, I offer an extended critical study of Sources. In chapter 7, I discuss Korsgaard’s formulation of what she calls “the normative question,” along with her rejection of moral realism as a candidate answer. In chapter 8, I discuss her autonomy-based answer to that question. I conclude with an Afterword, summing things up, and two appendices devoted to a discussion of Rawls’s views on moral realism and constructivism (on which Korsgaard heavily relies), and a detailed discussion of the critical reception of Korsgaard’s main argument, intended to identify lacunae in standard reconstructions of that argument.

Notes


CHAPTER 7:
KORSGAARD ON THE NORMATIVE QUESTION

7.1 Introduction

Among the many merits of Christine Korsgaard's *The Sources of Normativity* (hereafter *Sources*) is the book's self-conscious attention to what we may call the erotetics of ethics, that is, the methodological presuppositions that determine what questions we ask in the field. What, if anything (we may ask) is the single agenda-setting question for ethics, that is, the question whose answer sets the agenda for ethical inquiry? How do we generate such a question and verify its fundamentality? Why take the agenda-setting question to be $A$ rather than $B$? Once we have the right question in hand, how do we go about answering it? How do we generate the right menu of candidate answers, and how do we exclude non-candidates without begging the question?

Where much of the contemporary analytic literature glosses over such issues, taking them to be settled or perhaps too tedious to be worth discussing,¹ Korsgaard (following Prichard’s “Does Moral Philosophy Rest on a Mistake?”) devotes a good quarter or third of her book to a detailed analysis of them. In this chapter, I examine her account, focusing narrowly on how she formulates what she calls "the normative question," and how she sets up the preliminary features of the answer. I argue that Korsgaard's theory fails on both counts: her question is misformulated, and the various
conditions she places on a viable answer are illegitimate. If we see how and why Korsgaard's account goes wrong, as I hope to show, we should be in a better position to see what it would take to get things right.

In Section 7.2, I discuss Korsgaard's formulation of the normative question. In Section 7.3, I discuss two of the three conditions she imposes on a feasible answer to the question, the “address” and “identity” conditions. In Section 7.4, I discuss the ways in which she excludes candidate answers, focusing on the way she sets up her taxonomy of theories, and her exclusion of moral realism. The discussion in the present chapter sets the stage for my critique of Korsgaard’s autonomy-based argument for the foundations of ethics in the next chapter.

7.2 Erotetics of the Normative Question

After a brief prologue, Korsgaard opens Sources with an account of the specific version of the normative question that motivates her inquiry in the rest of the book. Her aim here is to formulate what I've just called "the agenda-setting question," in light of what she (correctly in my view) sees as a general failure to address such issues. "Since many moral philosophers have not addressed the question directly," Korsgaard writes, "it is not always clear what their answers are." Philosophers' "tendency to conflate the normative question with other questions often results in the normative question being blocked or ignored." Hence the need to get the question right and to do so in a precise way.

A close look at Sources suggests, however, that despite her self-consciousness about the proper formulation of the normative question, Korsgaard poses it in several
different ways throughout the text. This leads to several problems, all of them variations on a single basic theme: the equivocality and multiplicity of what is supposed to be a single, univocal question requiring a single, univocal answer.

First consider the difficulty of generating the list of questions ab initio from the text of the book. Ordinarily, we might expect to be able to track discussion of a book's guiding concern simply by looking at the index entries for the relevant phrase. There is, to be sure, an index entry in the book for "the normative question," but the entries refer to the question as though it were univocal when (as I'll argue below) it is not. In consequence, the reader interested in discussing Korsgaard's formulation of the question is forced to track and flag it on his own while working through the text. Depending on how one counts—and that is a controversial issue of its own—Korsgaard poses some seven distinct versions of the normative question in Sources, as follows:

1. Why do we need to use normative concepts, and what is their source?
2. Is there anything that beings like me (like us) are required to do, or "have to" or "must" do?
3. What is the status or basis of moral obligation?
4. Why is that I can be obliged/required to do ~X when desire bids me to do X?
5. When desire bids, should I follow its promptings or not?
6. What justifies the claims that morality makes on me (on us)?
7. What ends a regress of practical or moral justifications?

In one sense, the preceding questions are similar enough to justify their being thought of as one and the same "normative question": each of them clearly asks a fundamental question about the nature and status of moral or practical norms, and each answer must
bear some implicit relation to the others. In another sense, however, each question is, however subtly, sufficiently different from the others to justify our treating them as different questions:

(i) Question (1) asks about moral concepts; questions (2)-(7) do not.

(ii) Question (7) explicitly invokes the notion of a justificatory regress; questions (1)-(6) do not.

(iii) Questions (4)-(5) make essential reference to desire; the other questions do not.

(iv) Questions (3) and (6) make reference to morality; the other questions do not. 4

(v) Question (2) makes reference to acts and requirements but makes no reference to moral concepts, regresses, desire, or morality as do the other questions.

Since each of these questions invokes concepts not included in the other questions, they obviously aren't the same question. And since the relations among the various concepts invoked is far from clear—as are the relations among types of question and their levels of abstraction—it isn't clear how the questions are related to one another, either. At one level, this creates a purely expository problem. Since Korsgaard conflates the questions, she conflates the answers to them. There is thus, at a purely expository level, a persistent mismatch between questions and answers throughout the book. 5

But the deeper problem here is conceptual, not expository. It is not just that Korsgaard fails to distinguish between nominally different questions, but that she conflates substantially different questions without seeing the significance of doing so for her argument. There is first the conflation of what I'll call moral concept questions, such as (1), with questions about normative force, such as (2)-(7). There is secondly a
conflation within the category of normative force questions, between the *prescriptively presuppositionless* and the *prescriptively presupposition-laden* (my terminology). The first conflation, though confusing, is relatively minor. The latter conflation, I contend, is fatal to the theory.

7.2.1 The Moral-Concept Question and Internalism

Consider the difference between question (1) and the other six questions. Question (1) introduces the normative problem by way of remarks about “moral concepts.” As Korsgaard points out, we all use moral concepts, and one function of a "theory of moral concepts" is to justify our use of them by identifying their source. By contrast, questions (2)-(7) introduce a problem about the normative force of the moral claims that bind us: people make claims of various sorts of one another, and of themselves; questions (2)-(7) raise justificatory issues about why these claims really oblige us, and why we should hold ourselves responsible for living (or not living) up to them.

The moral concept issue is on the face of it distinct from the normative force issue. It is one question what the *source* of moral concepts is, another issue why we need to use them, and still another why we ought to regard ourselves as bound by the claims consisting of them. The first is an epistemic or semantic claim about the relation between concepts and the world; the second is a quasi-psychological question about our cognitive needs; and the third is a meta-ethical question about the basis of moral obligation. The three issues are no doubt related in some way, but the relationships among them are far
from obvious.Granted that moral claims consist of moral concepts in various semantic
arrangements, it doesn't follow from that that an answer to (1) answers any of (2)-(7).

One possible bridge between the moral concept and normative force issues might
be a version of motivational internalism. Korsgaard cursorily defines internalism in
Sources as “the view that moral judgments necessarily motivate” us but elsewhere
defines it more precisely as follows:

An internalist theory is a theory according to which the knowledge (or the truth of
the acceptance) of a moral judgment implies the existence of a motive (not
necessarily overriding) for acting on that judgment.6

The initial trouble, of course, is that so described, internalism, even if true, is about
judgments, not concepts, so that its connection to question (1) is obscure.

As far as the text of Sources is concerned, the matter ends there, but we could, as
a matter of sheer speculation, wonder a bit about what it would take to bridge the gap
between judgment internalism and (1). Drawing perhaps on the conception of moral
motivation at work in Socrates and Plato,7 we might imagine a view according to which
the sheer grasp by a moral agent of a normative concept motivates the agent’s actions—
or at least tends to—in a c-like way. Call this concept-internalism:

S’s grasp of c motivates (or has a tendency to motivate) S’s actions in a c-like way.

The adoption of such an internalism would narrow the gap between (1) and the other
versions of the normative question, as follows. Imagine an answer to (1), A, which gives
us an account of the source or basis of normative concepts. Now suppose we adopt
concept-internalism. In this case, A would give us the account in virtue of which S’s
grasp of c motivated S’s actions in a c-like way. Thus, for instance, on achieving a grasp
of the concept ‘morality’, $S$ would be motivated in the direction of morality as such. And that in turn would seem to answer many of the remaining questions; if the source of normative concepts was identical to the source of normativity, and the grasp of a concept ensured motivation, agents would, on grasping the relevant concepts, be motivated in the direction of the norms implicit in them. It would follow that the route to answering the normative question would proceed principally via a theory of normative concepts.

The trouble with this story, however, is that though Korsgaard in *Sources* and elsewhere endorses and partially defends judgment internalism, she doesn’t to my knowledge defend anything like concept-internalism. Indeed, truth or falsity aside, concept-internalism is a highly eccentric view. So the distinction between question (1) and the other questions is at best an unresolved and (at present) unresolvable matter.

Having identified this problem, however, I'll set it to one side, since it seems more to be a gap in Korsgaard's argument than a fatal objection. In any case, the rejection of (1) still leaves (2)-(7) untouched. The problem in the present context is not that concept-internalism is false, after all, but that Korsgaard needs it but leaves it inexplicit and undefended. But while the "moral concept" strain shows up from time to time in *Sources*—where it's consistently conflated with normative force questions—moral concept questions don't really drive Korsgaard's project in any important way. So in cases where we find references to both sorts of question in the text, I propose to take normative force questions as overriding moral concept questions.
7.2.2 The Problem of Normative Presupposition-Ladenness

Bracketing question (1), then, we're left with questions (2)-(7). As we've seen, while it's clear that these are all questions on a common theme, it's also clear that they're not the same question. Attentive to the differences between them, we may wonder how the differences matter to Korsgaard’s inquiry. Is one question more important or more fundamental than the others? Are some reducible to others? Must the questions be addressed in some particular order, or can they be addressed in any order?

Korsgaard herself gives us an indirect clue to answering these queries in her discussion in Sources of Prichard's "Does Moral Philosophy Rest on a Mistake?" But while she identifies some real problems, she does not, in my view, get at the deep problem that underlies them. "Discussions of normativity often founder,” she says, because of unexamined assumptions about the normatively loaded word [that the question is about]. There are two problems here. First, philosophers making different assumptions about which is the normatively loaded word may fail to understand each other. The second and perhaps most serious problem is that all of the ways of formulating the normative question that I have just mentioned suffer from the fact that they are readily confused with different questions…This tendency to conflate the normative question with other questions often results in the normative question being blocked or ignored....Part of what I have tried to do in this lecture is to raise the normative question in a way that is independent of our more ordinary normative concepts and words. 8

Korsgaard is right about both problems: the first is what I’ve previously called the “expository” problem, the second the “deep conceptual one,” and the second is more serious than the first. But Korsgaard misses some relevant and important points here. The first is that, as can be inferred from the list of questions in section 7.2 above, Korsgaard herself falls into the second confusion she describes here: she herself conflates questions (1)-(7) throughout Sources. As a result, I think, she also falls into the first confusion:
having conflated questions (1)-(7), which involve different normative concepts, she conflates the concepts as well, thereby conflating various assumptions about the normatively loaded word throughout the text. The deeper problem, however, is that while Korsgaard is sensitive to the need to distinguish the questions (however successful she ends up being at doing so), she seems relatively unworried about what she freely describes as the *loadedness* of the questions. But surely there is a problem lurking here. A loaded question covertly leads the questioner to an answer; the more loaded the question, the stronger the guarantee of a particular answer. In the limiting case, a question can be so loaded as to function as a merely rhetorical way of rephrasing an answer in the interrogative voice. But to ask ‘Why is p the case?’ as a way of asserting that *p is* the case surely begs the question: in this case, we might say, ‘Why *p*?’ is a question-begging question.

In light of this, it would be helpful, if possible, to arrange Korsgaard’s questions from least to most loaded. I’ll refer to the least/most loaded distinction as that between *normatively presuppositionless* and *normatively presupposition-laden* questions. But the distinction is itself a tricky one, and could use some elaboration.

In one sense, of course, there is no such thing as a literally presuppositionless question. A question is a request for information of a certain kind, where the request invariably involves, as a condition of its intelligibility, tacit assertion-conditions about the sort of information sought. Subtract the assertion-conditions, and one subtracts from the question the sort of information one is seeking; subtract enough, and one no longer has an intelligible question.
In some cases, the assertion-conditions are background assumptions not asserted in the question itself but presupposed by it as the necessary context for the question's intelligibility. The most obvious examples are one-word interrogatives: who?, what?, when?, where?, why?, and how? Any such one-word interrogative can function on its own as a question, but there is of course no way to know what such a question is asking absent information about the relevant context. In other cases, the assertion-conditions are in some way embedded in the wording of the question itself. The classic case, of course, is "Have you stopped beating your wife yet?"—where the tacit assumption, embedded into the very formulation of the question, is the claim that the person being questioned had once beaten his wife.

The relevant distinction between presuppositionless and presupposition-laden versions of the normative question, then, is not going to be one between versions that have and those that lack presuppositions *per se*, but rather those whose presuppositions or lack thereof are relevant to an inquiry into the foundations of ethics. In this context, the more relevant distinction would seem to be one between versions of a question that make no (or very few or relatively trivial) presuppositions about what the content of morality must (or must not) be, and versions of the question that make some such presuppositions. A normatively presuppositionless question, we might say, is one whose assertion-conditions contain no (or few or uncontroversial) normative claims, while a prescriptively presupposition-laden question is one whose assertion-conditions contain some (or many or controversial) normative claims, varying in degree of presupposition-ladenness by the number and/or controversiality of the normative claims embedded in or presupposed by the question.
While I think this distinction will do the trick for present purposes, we should note two complications before proceeding to apply it. First, reasonable disagreement can arise about the degrees of normative presupposition-ladenness in a given question. Consider the question "Why should I be moral?" as against the question "Why must I universalize the maxim of my action?" Both are clearly normatively presupposition-laden—the first presupposes that I should be moral, the second that I must universalize the maxim of my action. But it’s not obvious which is more so, or even how to go about counting the presuppositions with any accuracy.

Second, it is perhaps an odd truth but a truth nonetheless that no question is entirely normatively presuppositionless. Every question, being a request for information, presupposes the legitimacy of the practice of question-asking, a practice itself based on norms. And question-asking is itself a form of inquiry that depends for its possibility on the norms that guide it. If we define a norm minimally as a proposition that serves to guide voluntary action, then agents involved in erotetic activity ipso facto presuppose some norms even as they ask questions about the ultimate source of normativity. So normative presuppositionlessness may well turn out to be a matter of degree: a question may presuppose just the norms required to make sense of the activity of question-asking without presupposing very much more. (I’ll return to this issue in section 7.3.1.)

These complications aside, there is some distinction to be made, at least for purposes of differentiating degrees of presuppositionlessness and presupposition-ladenness on the preceding list of questions. How then to arrange Korsgaard’s questions in terms of their fundamentality (i.e., from least to most presupposition-laden)?
There are a couple of problems here. First, there are different ways of arranging the questions. Second, the moral/practical distinction keeps interfering.

The latter question has to be settled first. Do you want to distinguish moral and practical requirements, or not? Korsgaard isn’t clear.

Suppose you don’t distinguish. Then one possibility is to take (2) as basic and arrange all of the other questions in terms of it. The answer to (2) tells you that action is necessary. The necessity of action gets the regress started and raises (7) as a question. Bear in mind that (3) and (6) are related but different as are (4) and (5). The answer to (7) is *ipso facto* an answer to (3) and (6), and gives an answer to (4) and (5).

Suppose you distinguish. Then you have a complicated story depending on how moral and non-moral requirements relate to each other. One possibility is that moral overrides non-moral. A second possibility is that non-moral overrides moral. A third possibility is that neither overrides. A fourth possibility is that there is no single answer as to which overrides the other. In this case (2) remains basic and gives rise to (7). But (7) gives rise to two different sorts of normativity. Assume the fourth possibility *ex hypothesi*. Then there are cases in which moral reasons override non-moral, but not always. In that case, some aspect of the answer to (7) will answer (3)-(6).

The relevant issue is that on either interpretation (2) has a fundamentality in virtue of its normative presuppositionlessness that the other questions lack. It simply asks whether an agent has to act and why, placing no stipulations on the form of the answer. The answer could take any form and have any content. An answer consisting of hypothetical imperatives would answer the question as directly as an answer consisting of categorical ones. A consequentialist answer would answer the question as directly as a
non-consequentialist. Etc. Crucially, the question genuinely leaves open the possibility that there are no practical requirements at all: “Are there X’s?” doesn’t presuppose the existence of X’s. The question asks whether there are, within a context in which no answer is taken for granted. It’s precisely whether there are that must be answered, and answered before moving to the next set of questions.

Compare (2) with (6):

(6) What justifies the claims that morality makes on me?

Question (2) asks whether we should act, without assuming that we should. Question (6), by contrast, presupposes all the information that (2) seeks, and more besides. It not only assumes that we are required to act, but that “morality” makes claims on us, and that these claims are at least prima facie justified. As Korsgaard later remarks in passing, since she assumes that moral requirements are categorical in force, (6) gives moral claims a specific normative force and content.

We’re now in a position to see the trouble with Korsgaard’s conflation of normative force questions. Questions like (2) are broad. Questions like (6) are narrow. And there are questions intermediate between them. If we conflate (2)-like questions with (6)-like questions, we lose our sense of the scope of the question we’re asking. What we lose, in particular, is a sense of the order in which the questions must be answered. If the questions are answered in turn, we move from broad to narrow questions, using the answers to the broad questions as a basis for answering the progressively narrower ones. But if the procedure is reversed, three fatal problems arise.

The first is that if the broad and narrow questions are conflated, and only the narrow ones are answered, the broad questions go unanswered—in which case, the less
fundamental questions are answered at the expense of the most fundamental ones. And so we learn that we are obliged to act on categorical imperatives without learning why we are obliged to act.

The second is that if the broad and narrow questions are conflated, and the narrow ones presuppose answers to the broad ones, then an answer to the narrow question can give the illusion that the broad question has already been answered. Thus if we’re told that we have reason to be moral without being told why we’re obliged to act, the supposed ‘reason to be moral’ can give the illusion that the question ‘why act?’ has been answered when it has not.

The third is that if the answers to the broad and narrow questions are different, then the conflation of the broad and narrow questions can conceal that fact. With the answer to a narrow question in hand, we can easily forget that no answer to the broad question was given; having forgotten that, we will miss the ways in which the answer to the narrow question piggybacks covertly on the answer to the broad question—without ever treating it as a separate question, and even when the two answers are incompatible with each other. And so a Kantian account of the categorical imperative can depend for its content on, say, an Aristotelian conception of flourishing incompatible with the Kantian account.

This, in my view, is exactly what happens in Sources. Korsgaard poses questions as broad as (2) and as narrow as (6). She conflates all of them, but directly answers only the narrowest of them. Having done so, she tacitly presupposes answers to the broader questions. The answer she ultimately gives is a Kantian deontological one. This Kantian account is flatly incompatible with the Aristotelian thesis that the requirements of self-
interest or flourishing furnish answers to the broad questions (and by implication the narrow ones). Korsgaard never directly addresses the broadest question, question (2), but she does address the narrowest question, question (6). In conflating the two questions, she naturally conflates the answers, and so the answer to question (6) implicitly is taken as an answer to question (2). The result is that Korsgaard offers a defense of the categorical imperative that tells us *what it is that we must do* without ever explaining *why there is anything we must do*. The answer to the latter query is simply assumed but used as a tacit basis for the former.

For now, however, we needn’t determine whether the tacit answer to (2) is in fact incompatible with Korsgaard’s defense of the categorical imperative. The conflation of the versions of the normative question is a problem of its own, and worth identifying in its own right.

7.3 Conditions on the Answer to the Normative Question

In the previous section, I discussed the presuppositions tacitly embedded in Korsgaard's versions of the normative question, and noted how they illicitly lead to the answer that Korsgaard wants to give. But Korsgaard also offers a series of *explicit* conditions on the answer. "To be successful," she says, "there are three conditions which the answer must meet: it must actually succeed in addressing someone in the first-person position of the agent who demands a justification of the claims that morality makes on him"; it must in a certain sense be "transparent"; and it "must appeal, in a deep way, to our sense of who we are, to our sense of identity." These somewhat innocuous-looking conditions, I contend, give rise to the same sorts of problems as do the presuppositions
built into questions (2)-(7) in the previous section. I discuss the address and identity conditions below, leaving the transparency condition aside as essentially unproblematic.

7.3.1 The Address Condition

The first of Korsgaard's conditions is what she calls the address condition:

First, the answer must actually succeed in addressing someone in that position. It must not merely specify what we might say, in the third person, about an agent who challenges or ignores the existence of moral claims. Every moral theory defines its concepts in a way that allows us to say something negative about people who do that—say, they are amoral or bad. But an agent who doubts whether he must really do what morality says also doubts whether it's so bad to be morally bad, so the bare possibility of this sort of criticism settles nothing.⁹

There's certainly something right about Korsgaard's address condition: agents can ask questions from the first-person perspective, and an answer ought to be responsive to the question actually asked; the answer ought, in the lawyer's phrase, to be an "affirmative response" to the question. But Korsgaard's formulation of the condition is stronger than this, and so, more problematic.

The address condition is intended to rule out some possibilities and mandate others. What is radically unclear, however, is how it does so, i.e., what counts, in practice, as the sort of answer that the condition rules in and the sort it rules out. If we take the condition literally, it seems merely to be making a grammatical point: answer a first-personal question with a first-personal answer, not a third-personal. But construed this way, the condition is trivial: any third-personal answer to a given question can be transformed into a first-personal one by the simple expedient of being rephrased, in which case the condition is toothless. If so, the condition lacks content.
Consider an example. In the Platonic dialogues, for instance, Socrates frequently tries to convince his skeptical interlocutors that morality promotes a human being's highest interests, or putting it the other way around, that immoral action is self-subverting. So in the Republic, Glaucon famously makes the following challenge:

But I've yet to hear anyone defend justice [dike] in the way I want, proving that it is better than injustice. I want to hear it praised by itself....

Phrased as a first-personal question, Glaucon is asking

Why should I be just?

where "just" translates dike, which arguably could be taken as a synonym for "what's right." When Socrates finally gets around to answering that question some seven books later, one of the three answers he gives is:

[T]he best, the most just, and the most happy [person] is the most kingly [one], who rules like a king over himself and...the worst, the most unjust and the most wretched [person] is the most tyrannical [one], who most tyrannizes himself and the city he rules.

Phrased less metaphorically, Socrates is responding to Glaucon's first-person question in the third-person:

One should be just because one is better off if one is, and worse off to the extent that one is not.

So Socrates gives a third-personal response to a first-personal question, and goes on to describe in some detail the delights of the just life in contrast to the wretchedness of the unjust one. Whatever one thinks of the plausibility of Socrates' claim, the problem with it cannot be a failure to respond to or address Glaucon's question: it's as direct a response as Glaucon could expect. And if Glaucon finds it insufficiently direct, he need merely
change the third-personal references to first-personal ones, and that should be the end of
the story. But if so, the address condition is doing no substantive work.

Consider yet another example of the same point: the beginning of Mill's proof of
the principle of utility in *Utilitarianism.*

The only proof capable of being given that an object is visible, is that people
actually see it. The only proof that a sound is audible, is that people actually hear
it: and so of the other sources of our experience. In like manner, I apprehend, the
sole evidence it is possible to produce that anything is desirable is that people do
actually desire it…No reason can be given why the general happiness is desirable,
except that each person, so far as he believes it attainable, desires his own
happiness.¹⁰

The proof here is stated in the third-person: what is visible, audible, desirable. But it
refers to facts only accessible from a first-person perspective: one can only grasp the
meaning of the proof from a first-person perspective: what *I* see, hear, desire.

So does Mill's proof "address" someone asking Korsgaard's normative question?
Korsgaard insists that it does not,¹¹ but her argument on the point is unclear and
unconvincing. One oddity of the discussion is that despite the length of the discussion,
Korsgaard does not directly discuss Mill's proof at all, except for one very brief and (in
this context) uninformative comment. What she discusses is (not the proof per se but)
what she takes to follow from the fact that the proof commits Mill to a form of
externalism about motivation. But that misses the feature of the proof to which I've just
alluded: externalist or not, the meaning of the proof can only be gotten from a first-person
perspective. The difficulty here is that what Korsgaard takes to be a paradigm case of not
addressing the question seems in fact to be a paradigm case of addressing it.

Now, Korsgaard might at this point respond that, Plato and Mill aside, the address
condition is more-than-merely grammatical: some third-personal answers simply aren't
translatable into a first-person perspective. They merely state third-personal facts about the agent. An agent could, we might imagine, listen to such an answer, translate it into the first-person, and yet legitimately claim to be unsatisfied by the translation on the grounds that it simply left him cold. By his lights, the answer merely says something about him from the third-personal perspective, a fact that remains unchanged when the answer is rephrased into the first-person. Since he "doubts whether it's so bad to be morally bad," he will naturally doubt the relevance to him of an answer that says that it is. He might, ex hypothesi dig in his heels and insist that neither Socrates' answer to Glaucon nor Mill's proof in fact met the address condition in his case.

This possibility would seem to push us back to the first issue concerning the distinction between questions and questioners. The obvious question to ask is why the challenger's reaction should have any bearing at all on the adequacy of the answer—e.g., Socrates' or Mill's.

There is first Korsgaard's conflation of responsiveness to the question with responsiveness to the questioner, and generally, of semantic with psychological or epistemic issues. The question, we are told, must "address" the questioner from his perspective. But in conflating responsiveness to the question and responsiveness to the questioner, Korsgaard excludes answers that fail to "address" the questioner through defects that are present not in the answer but in the questioner.12

Take a question $Q$ and an answer $A$, asked of a subject $S$. The following six possibilities ought to be distinguished:

1. $A_1$ answers $Q$ and convinces $S$.
2. $A_2$ answers $Q$, but doesn't convince $S$.
3. $A_3$ doesn't answer $Q$, and (so) doesn't convince $S$.  

278
4. $A_4$ doesn't answer $Q$ but somehow manages to convince $S$ anyway.
5. $A_5$ answers $Q$ but, while not convincing $S$, puts $S$ sufficiently on the defensive to rebut the skepticism that led $S$ to pose the challenge contained in $Q$.
6. $A_6$ doesn't answer $Q$, but while not convincing $S$, puts $S$ sufficiently on the defensive to rebut the skepticism that led $S$ to pose the challenge contained in $Q$.

$A_1$ and $A_4$ satisfy the address condition, but it's not clear why their doing so makes them any better as answers to the question than $A_2$ or $A_5$. Consider them in pairwise fashion.

Both $A_1$ and $A_2$ answer $Q$, but only $A_1$ convinces $S$. What if $A_2$'s failure to convince $S$ arises from moral or epistemic failure on $S$'s part? In that case, the address condition tells us nothing about the adequacy of the answer and merely appeases the moral or epistemic defects of the questioner. Again, $A_4$ meets the address condition to a higher degree than does $A_5$: it has a more drastic doxastic effect on $S$. But $A_4$'s having a more drastic effect may be explainable by the slick, gimmicky, or glib fashion in which it is expressed, or by $S$'s credulity, or by a combination of both factors. Again, the address condition tells us nothing about the adequacy of the answer; indeed, it masks the inadequacy of some answers and the defects of some audiences.

Korsgaard might respond that the address condition is after all just a necessary condition on an adequate answer, and that further conditions might be required to weed out further inadequacies. But the fact remains that the address condition remains problematic even as a necessary condition. Taken literally, the condition constitutes an insuperable bar even to a dialectical demonstration of the Law of Noncontradiction (LNC). If I stubbornly refuse to accept the LNC, I might well insist that you convince me of it by wielding the address condition against any attempt. “The LNC simply does not
speak to me,” I might insist. But my insistence would not change the dialectical situation: my justificatory request is unreasonable.

Korsgaard herself stipulates that the question must respond to some but not all agents. "For this exercise to work," she writes, referring to the exercise of answering the normative question, “we have to eliminate [certain] possibilities, and imagine that [the] agent is sincere and reasonable, and does really want to know." Call this sort of person radically disaffected. So the normative question need not address the insincere and/or unreasonable at all, but it must address the sincere, reasonable inquirer only from the first-person perspective, not the third.

There is a question-begging element in Korsgaard's stipulation that the address condition exclude the "unreasonable." And this, too, leads to a dilemma. Either reasonability is defined in terms of a propensity to accept the norms of practical reason or not. If it is, then the reasonability stipulation begs the question: the address condition requires the agent to have a prior propensity to accept the norms of practical reason, in which case the agent is posing the normative question while having a prior propensity to accept the answer to it. If not, then the agent must accept the answer to the normative question on some basis other than a prior propensity to accept the norms of practical reason. It's unclear what this could be, but one possibility is that the agent has a commitment to epistemic norms but not to practical-but-not-epistemic ones. This possibility has some plausibility, but it is one that Korsgaard explicitly rejects.

Korsgaard's stipulation brings me to my third objection to the address condition: its unmotivated, ad hoc quality. Consider the insincere and unreasonable agents whom Korsgaard excludes. Suppose one of them complains: "You clearly aren't addressing me
in your answer to the normative question. So your answer is deficient.” The hypothetical Korsgaardian response to this objection would have to be either first-personal or third-personal. Either way, it runs into trouble.

Suppose that it's stated first-personally. In this case, the answer seems impossible to generate: how does one convince an insincere or unreasonable person, qua insincere or unreasonable, of anything to his first-personal satisfaction? Whatever one says is guaranteed to fail.

Suppose then that it's stated third-personally. The problem now becomes that if the whole problem with third-personal answers was that they run the danger of not "addressing" the agents to whom they're addressed, that would apply as well to Korsgaard's answer to the insincere (etc.) agent who asks why he isn't being included in the game.

Korsgaard might well respond that that's hardly a "danger" at all; we're not obliged to address the insincere or unreasonable agent, since there is a sense in which they either don't want to be addressed or can't be. Perhaps, but if so, shouldn't the very same thing apply to the radically disaffected but sincere agent? Why are we required to produce an answer that satisfied the disaffected-but-sincere agent from the first-person perspective if in the same breath we can stipulatively exclude others for reasons that cannot satisfy them from a first-person perspective? In other words, if the answer to the normative question need not take stock of the insincere or unreasonable at all, why must it take stock of the disaffected in such a way as to satisfy them from a first-person perspective? Why aren't the insincere, the unreasonable, and the disaffected all in the same boat?
The dilemma here is this: either the address condition is to be applied without exclusionary stipulations or with them. If not, it can't be satisfied because the insincere agent need never regard himself as addressed. If so, then either the stipulations have a rationale or not. If not, the stipulations are ad hoc and so is the address condition. If so, either the rationale for the stipulations meets the pure address condition or not. Obviously, the rationale for the stipulations cannot meet the address condition sans stipulations for reasons given above. So the rationale for the stipulations on the address condition are themselves not subject to the address conditions. But if that is so, we can legitimately ask why the address condition should be taken seriously.

7.3.2 The Identity Condition

I move now to the third of Korsgaard's conditions, namely, that the answer to the normative question must appeal to our sense of identity:

As I have been emphasizing, morality can ask hard things of us, sometimes even that we should be prepared to sacrifice our lives in its name. This places a demanding condition on a successful answer to the normative question: it must show that sometimes doing the wrong things is as bad or worse than [sic] death. And for most human beings on most occasions, the only thing that could be as bad or worse than [sic] death is something that for us amounts to death—not being ourselves any more....If moral claims are ever worth dying for, then violating them must be, in a similar way, worse than death. And this means that they must issue in a deep way from our sense of who we are.

Later in Sources and elsewhere in her writings, Korsgaard elaborates on this conception of identity, calling it "practical identity": a practical identity, she writes, is "a description [of yourself] 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 identity so
understood is simultaneously the basis of obligation, integrity, and self-respect; once adopted, we identify with our practical identity in such a way that, in violating it, we feel as though we are violating ourselves, undermining our own self-constructed identity.

Korsgaard's conception of practical identity is both controversial and interesting in its own right, but my concern here is with identity as an adequacy-condition on her answer to the normative question. Can Korsgaard legitimately raise practical identity at this stage of her inquiry, effectively building it into the question itself? I think the answer is "no," for four interlocking reasons, each related to the other, and each related to the issues I've previously raised.

The first and most fundamental problem is that Korsgaard has simply assumed without argument that each of us has or ought to cultivate a practical identity. This assumes, at a minimum, that the need for action is a given, that we do value ourselves (and do so in the relevant way), and that we find our lives worth undertaking. In taking these things for granted, Korsgaard not only leaves the prescriptively presuppositionless version of the normative question unanswered, but packs a lot of unargued content into any candidate answer.

A second problem arises from the first. Korsgaard's address condition had stipulated that the answer to the normative question should answer a skeptic from a first-personal perspective. But we may legitimately wonder what Korsgaard would say to the skeptic who questions whether he needs a practical identity. Once again, Korsgaard faces a dilemma here. On the one hand, her answer to the normative question cannot satisfy such a skeptic, precisely because it assumes without argument that we need a practical identity (and states that assumption, incidentally, in the third person). On the other hand,
the address condition tells us that the normative question must be able to satisfy virtually any sort of skeptic, no matter how radically disaffected—and do so to his first-personal satisfaction—so long as that skeptic is sincere and looking for an answer. Faced with a sincere and reasonable skeptic about Korsgaardian practical identity, the two conditions will inevitably contradict each other.

A third problem arises from an ambiguity in the concept of practical identity itself. Even if we could assume that we all need a practical identity, that by itself would not tell us what practical identity to get, or what constraints there are on permissible practical identities. Practical identities are obviously going to vary with the persons whose identities they are. Consider just the differences in practical identity between you, me, Osama bin Laden, Nelson Mandela, Crazy Horse, Joseph Stalin, Queen Elizabeth I, the Emperor Shah Jahan, Alcibiades, Jesus, Pele, and Steven Spielberg. If *ex hypothesi*, each of these persons has a practical identity, then to say that the normative answer must appeal to each one's practical identity is either to make the task of answering impossible (due to the sheer breadth and variety of identities to encompass, and obvious incompatibilities between them) or the answer vacuous (consisting in the ridiculously thin normative content that all those identities have in common, on the highly unlikely possibility that such a thing exists). To avoid this result, Korsgaard must be restricting the scope of permissible practical identities to some proper subset of the human population: her appeal to "our sense of who we are" must refer only to *some* of us. But now we have a normative question that begs the question—in other words, a question fashioned in such a way that asking it leads straight to Korsgaard's answer.
A final point. Korsgaard's remarks about identity seem to suggest that morality can require sacrifice of one's life because immoral action would disfigure one's identity and such disfigurement is worse than death. Besides being controversial, the conception of disfigurement at issue here is highly metaphorical. The analogy is with physical disfigurement of the body: just as one would (or ought?) to want to die rather than undergo some horrible torture, so one would (or should) want to die rather than undergo the torture associated with a self-disfigured practical identity. In the physical case, death is preferable to torture either because the pain of torture is unbearable or because life is only worth living given a minimum of functioning that torture makes impossible even after it's over.

These are large (and somewhat ambiguous) claims, but let's assume that they're right. What's striking about them is their apparent attachment to a form of egoism. If the metaphor of disfigurement means anything, it means that as physical torture is bad for the body, moral disfigurement is bad for the person. If the normative answer must show us that violations of moral norms disfigure us, an answer would surely fail if it showed us that such violations not only didn't disfigure us but redounded instead to our psychological health. Admittedly, Korsgaard doesn't quite say that every violation of any moral norm always disfigures us; that would overstate things. Still, if the identity condition has any content at all, it has to rule out a fundamentally self-alienating conception of morality, and rule in one that at some basic level conduces to psychological health. So while it might go too far to say that the identity condition leads us straight to a form of ethical egoism, it seems reasonable to say that the condition lends an egoistic bias to any candidate answer to the normative question.
There are, I think, two problems here. The first, of course, is that the egoism is covertly packed into the formulation of the question. The other, however, is that Korsgaard vehemently insists that her account is not egoistic—indeed, insists here and elsewhere that egoism is an incoherent view and a "myth." For now, it's sufficient to say that if the identity condition entails egoism, as it \textit{prima facie} does, it is inconsistent with Korsgaard's larger theory.

7.4 Constricting the Range of Answers

I turn now from the way Korsgaard sets up the question to the way she sets up the answer. Schematically, we can think of her argument here as an argument by elimination. In section 1.2.3, Korsgaard identifies four candidate answers to the normative question—Voluntarism, Realism, Reflective Endorsement, and the Appeal to Autonomy (hereafter, "Autonomy"). The rest of Lecture 1 consists of critiques of Voluntarism and Realism intended to exclude them from the running. Lecture 2 offers a discussion of Reflective Endorsement that more or less takes it out of the running, as well. Lectures 3 and 4 then go on to defend what Korsgaard calls an "updated version" of the appeal to autonomy.

In this section, I offer a two-part critique of this part of Korsgaard's argument. I begin by taking issue with her menu of candidate answers to the normative question. I query what seems to me to be the most problematic—and in a sense, the most severe—of Korsgaard's exclusions, her critique of moral realism. As we'll see, this critique does not succeed, and its failure pushes Korsgaard much too quickly from the normative question to her preferred answer. It also confuses rather than clarifies the nature of other possible answers.
7.4.1 Korsgaard’s Menu of Answers

Lecture 1 of *Sources* discusses and rejects Voluntarism and Realism as candidate answers to the normative question. As preface to examining what Korsgaard has to say about those theories, however, it is worth pausing a moment to comment on a meta-theoretical point that is easy to miss, namely, the adequacy of the taxonomy of theories she offers for discussion. Recall that her argument proceeds by elimination. So as a matter of structure, the argument begins by offering four candidate answers to the normative question, excludes the unsuccessful answers, and leaves us with Korsgaard's own answer as the "winner." One obvious pitfall in this mode of argument is the possibility that deficiencies in the formulation of the original list of candidates will skew or rig the outcome of the argument as a whole. Arguably, that is precisely what happens here. What strikes—or ought to strike—the reader is the unclarity and counter-intuitiveness of Korsgaard's taxonomy of theories, a fact that in my view explains many of the particular difficulties and obscurities of her discussion.¹⁵

Korsgaard defines Voluntarism as the view that moral "obligation derives from the command of someone [beside the agent] who has legitimate authority over the agent and so can make laws for her." She defines Realism as the view that "moral claims are normative if they are true, and true if there are intrinsically normative entities or facts which they correctly describe." Reflective Endorsement is the view that "morality is grounded in human nature" and so "is good for us." Finally, Autonomy is the view "that the source of the normativity of moral claims must be found in the agent's own will, in particular in the fact that the laws of morality are the laws of the agent's own will and that
its claims are ones she is prepared to make on herself." So described, this taxonomy suffers from several significant flaws. For one thing, not being mutually exclusive, its categories are prone to equivocation. Second, not being jointly exhaustive, the categories artificially constrict the range of possible alternatives Korsgaard is tasked with examining, and push the reader too quickly to her preferred option.

7.4.1.1 Non-Exclusivity of the Taxonomy

Korsgaard defines the theories in such a way as to make Realism, Reflective Endorsement and Autonomy three distinct theories, the first two of which fail to answer the normative question. But as she defines them, there is no good reason to think that Realism excludes either Reflective Endorsement or Autonomy. If Realism is compatible with either or both, Korsgaard's taxonomy fails to draw the intended contrast between them.

Realism is compatible with Reflective Endorsement just in case moral claims are normative because true (satisfying the first conjunct of the definition of Realism), while simultaneously being based on facts about human nature that are good for us (satisfying Reflective Endorsement), where the recognition of those facts by a rational agent was intrinsically motivating (satisfying the second conjunct of the definition of Realism). Consider some claims from Spinoza's *Ethics* that perfectly fit the bill:

P20. The more one strives, and is able, to seek his own advantage, that is, to preserve his own being, the more he is endowed with virtue; conversely, insofar as each one neglects his own advantage, that is, neglects to preserve his being, he lacks power.
Is Spinoza therefore a Realist or a Reflective Endorsement theorist? The distinction seems pointless in cases like this, and thus talks past them.

Korsgaard claims that what differentiates Realism from Reflective Endorsement is that for the Realist, moral claims are true, whereas for the Reflective Endorsement theorist, "The question is not 'are these claims true?' as it is for the realist. The reasons sought here are practical reasons; the idea is to show that morality is good for us." This formulation is ambiguous between saying that the Reflective Endorsement theorist rejects the bivalence of moral propositions, and saying that bivalence is simply not her primary theoretical concern. If the Reflective Endorsement theorist accepts bivalence without its being her primary concern, there is little but a difference in emphasis between Realism and Reflective Endorsement. But if Reflective Endorsement rejects bivalence, it's saying something rather mysterious, namely, that morality is "grounded" in human nature and good for us, but that the propositions identifying these facts are neither true nor false. So the following proposition is neither true nor false:

The virtue of honesty is grounded in human nature and its practice is good for the agent.

It simply is not clear what it means to say that the preceding claim functions as a "practical reason," "shows" us something, but doesn't take a truth-value.

Having seen that Realism might be compatible with Reflective Endorsement, we can easily see how it might also be compatible with Autonomy. All we have to do is imagine a theory—or find an example of one—that combines a commitment to autonomy on a Realist basis. Couldn't morality's claims be true and intrinsically motivating, while
making essential reference to autonomy? It's hard to see why not. Consider the following proposition:

*Ceteris paribus*, it is the case that the exercise of autonomous agency is grounded in the nature of human reason, is therefore good for agents.

This satisfies Realism, Reflective Endorsement, and Autonomy simultaneously. What then is the point of distinguishing between them?

7.4.1.2 Non-Exhaustiveness of the Taxonomy

The non-exhaustiveness of Korsgaard's taxonomy is in a sense just the mirror-image of the non-exclusivity of its categories. The non-exclusivity problem tells us that while Korsgaard takes each category to exclude the others, in fact some of the categories include the others. The exhaustiveness problem tells us that Korsgaard's taxonomy leaves no conceptual room for certain theories. The most obvious such theories of this sort are ones that combine and integrate precisely the commitments that Korsgaard regards as distinct. A paradigm in this respect is Aristotle's ethics, and even more so, neo-Aristotelian or perfectionist theories of a variety that combine a commitment to moral realism with a commitment to autonomy.

Korsgaard's taxonomy depends on the assumption that Realism is opposed to Reflective Endorsement and Autonomy, which implies that no significant theory could endorse all three at once. There are two ways of "endorsing" the three commitments. The simplest way would be to endorse their conjunction, full stop. A more sophisticated way would be to endorse the conjunction within the context of a coherent account of why the conjunction obtains. (Coherence comes in degrees, so that a theory might in this second
sense endorse the three commitments to higher or lesser degree, depending on the degree of coherence achieved by the account of why the conjunction obtains.) Arguably, Aristotle's ethics is committed to the simple-conjunction view while certain neo-Aristotelian theories are committed to the coherent-conjunction view.

Consider Aristotle's view. By Korsgaard's definition, he is committed to Realism: he takes moral claims to be true and takes *eudaimonia* to be intrinsically normative, at least on one interpretation of that phrase. But by her definition, he is also a Reflective Endorsement theorist: he takes virtue to be grounded in human nature and good for us. And surprisingly enough, Aristotle could even be seen as committed to Autonomy; he at least mentions autonomy in the *Nicomachean Ethics* and scholars have (perhaps controversially) attributed some such commitment to him.

One could, to be sure, justifiably quarrel with an autonomy-based interpretation of Aristotle's ethics. What one finds in the *Ethics* is just the barest commitment to claims about autonomy, and while there are resources in the texts for constructing an account more robustly inclusive of autonomy (e.g., *NE* IX), Aristotle himself does not provide one. But contemporary neo-Aristotelian philosophers have gone out of their way to do so, and in so doing, have explicitly sought to explain why an Aristotelian view naturally coheres with a commitment to autonomy. In failing to make room for such a theory, Korsgaard too easily dismisses an answer to the normative question that competes with her own.
7.4.2 The Exclusion of Moral Realism

I turn at last to Korsgaard's critique of Realism. The critique follows on Korsgaard's discussion of Voluntarism, which she rejects for fairly standard reasons. Realism then comes in as a response to Voluntarism. Whereas Voluntarism locates normativity in an authoritative agent of some kind, Realism locates it instead in the world. Recall the definitions mentioned above. Thus Voluntarism holds that moral "obligation derives from the command of someone [beside the agent] who has legitimate authority over the agent and so can make laws for her." By contrast, Realism holds that "moral claims are normative if they are true, and true if there are intrinsically normative entities that they correctly describe."

Korsgaard's argument against Realism may be summarized as follows. In attempting to locate normativity in the world, Realism makes normativity problematically "irreducible." Given this, Realism not only cannot answer the normative question but refuses to address it at all: it rejects the legitimacy of the question and \textit{ipso facto} defaults on an answer. Taking normativity to be self-evident, realists merely content themselves with writing and speaking in ways that are intended to increase confidence in realist ways of thinking about ethics. In an attempt to clarify these points, Korsgaard likens realism's defects with those of the cosmological argument (CA): "The realist like the cosmologist places the necessity where he wanted to find it. And the argument cannot even get started, unless you assume that there are some actions which it is necessary to do."

Having criticized Realism so heavily, however, Korsgaard makes something of an about-face at section 1.4.4 in order to salvage a commitment to what she takes to be a minimal form of realism. To this end, she distinguishes between what she calls
procedural realism (PR) and substantive realism (SR), affirming the first and rejecting the second. Presumably, then, the criticisms she makes of realism tout court apply to SR but not to PR, so that in rejecting SR and affirming PR, Korsgaard takes herself to be entitled to enjoy the benefits of a form of realism immune to its problems.

There are, as I see it, several interconnected difficulties with Korsgaard's critique of realism, all of them turning on the essential eccentricity of her account of it.

1. Lack of independent rationale. There is first the problem that in defining Realism in relation to Voluntarism—or as part of a dialectic that makes essential reference to Voluntarism—Korsgaard never gives Realism an independent motivation, and so never clarifies why (Voluntarism aside) anyone might adopt Realism in the first place. It is thus a mystery what problem Realism was intended to solve. Considering that Korsgaard's definition of realism differs markedly from many other definitions, it is also a mystery how we are to adjudicate between competing definitions of the term. If realism_K differs from other realisms, even if Korsgaard's critique of realism_K is successful, why think that it has any bearing on other forms of realism defined independently of Voluntarism?

2. Irreducibility. As we've seen, Korsgaard's basic critique of realism turns on what she calls its irreducibility. A first difficulty we face in evaluating this argument is to get clear on what Korsgaard means by irreducibility. As I read Sources, the text is ambiguous as between two very different claims. On the one hand, Korsgaard seems to be endorsing what I'll call a regress-rejection thesis about realism:

   Regress Rejection (RR): Realism rejects the normative question because it regards the regress problem as entirely spurious. On the standard view of practical regresses of justification, for any action A that is prima facie obligatory for S, we
can ask why it is obligatory, getting a further answer $B$, and then raising the same question with respect to $B$. The resulting chain is \{circular or \{if linear, then (finite or infinite)\}\}; since it can't be circular or infinite, it must be finite. On RR, the realist just adamantly refuses to ask 'why?' in the first instance and so refuses to get the regress going; the only permissible questions are whether $A$ belongs to the set of (real) moral obligations, and if so, whether $S$ is in circumstances that require performing $A$.

On the other hand, she also seems to be endorsing two possible versions of the premature terminus thesis about realism.

Premature Terminus (PT1): Realism doesn't so much reject the normative question or the regress problem as give the wrong answer by stopping the regress at a manifestly premature location. So for any action $A$ that is \textit{prima facie} obligatory for $S$, we can ask why it is. Unfortunately, the realist—though admitting the legitimacy of this "why?"—merely tells us that the chain ends precisely where it began, viz., with $A$.

Premature Terminus (PT 2): Realism doesn't so much reject the normative question or the regress problem as give the wrong answer by stopping the regress at a manifestly premature location. So for any action $A$ that is \textit{prima facie} obligatory for $S$, we can ask why it is. Unfortunately, the realist--though admitting the legitimacy of this "why"--simply refuses to take the chain out a sufficient number of links, stopping, say, at $B$ or $C$ when it's obvious that we haven't yet reached the terminus.

Obviously, RR is incompatible with either version of PT. But Korsgaard says things that support both interpretations.

In any case, I think RR can easily be dispensed with. Korsgaard (plausibly) ascribes RR to Clarke, Price, Prichard, Ross, Moore, and Nagel, but the view held in common by this group is better described as an intuitionist species of Realism, not as Realism per se. So to the extent that Kosgaard takes RR as her target, her target is much narrower than realism as such, and it's hard to see how anything momentous could turn
on accepting or rejecting it. (Note the conflation of epistemic and meta-ethical regresses here.)

That takes us to PT. I’ve distinguished two versions of this interpretation, simply to highlight an ambiguity in the text, but it will suffice for present purposes to focus on the second of the two. Here, I need to quote Korsgaard at some length:

As these arguments show, realism is a metaphysical position in the exact sense criticized by Kant. We can keep asking, ‘Why must I do what is right?’—Because it is commanded by God. –But why must I do what is commanded by God? –and so on, in a way that apparently can go on forever. This is what Kant called a search for the unconditioned—in this case, for something which will bring the reiteration of ‘but why must I do that?’ to an end. The unconditional answer must be one that makes it impossible, unnecessary or incoherent to ask why again. The realist move is to bring this regress to an end by fiat: he declares that some things are intrinsically normative.…The very nature of these intrinsically normative entities is supposed to forbid further questioning. Having discovered that he needs an unconditional answer, the realist straightaway concludes that he has found one….A comparison will help to show why this is metaphysical. Consider the cosmological argument for the existence of God, which purports to prove God’s existence by proving that there must be a necessary being.…As Hume pointed out in his Dialogues Concerning Natural Religion, there are two problems here. First of all, so far as the argument goes, anything could be the necessary being. It could be matter, or the universe, or the sun.…And second, unless you assume that even contingent beings must in some sense by necessary—that is, that there must be an explanation which shows that they must have existed—the argument cannot even get started.

Having discovered that obligation cannot exist unless there are actions which it is necessary to do, the realist concludes that there are such actions, and that they are the very ones we have always thought were necessary, the traditional moral duties.…The realist like the cosmologist places the necessity where he wanted to find it. And the argument cannot even get started, unless you assume that there are some actions which it is necessary to do.

The difficulty here is plain. The metaphysical view that intrinsically normative entities or properties exist must be supported by our confidence that we really do have obligations. It is because we are confident that obligation is real that we are prepared to believe in the existence of some sort of objective values. But for that very reason the appeal to the existence of objective values cannot be used to support our confidence. And the normative question arises when our confidence
has been shaken, whether by philosophy or by the exigencies of life. So realism cannot answer the normative question.

There are in fact two distinct criticisms here. The first, less important one, is that as the CA is indeterminate, so is Realism: anything could end the CA’s regress (not necessarily God); so anything could end the realist’s regress (not necessarily intrinsically normative entities). The second one is that as the CA begs the question, so too does Realism: the CA only gets off the ground because the cosmologist assumes that the worlds needs an explanation in a necessary being; Realism only gets off the ground because the Realist assumes ab initio that we have obligations that have to be accounted-for by a sort of Ur-obligation. The cosmologist never explains why the world needs an explanation; the Realist never explains why we have obligations ab initio. Hence, Korsgaard concludes, Realism fails as does the CA. Setting aside the cogency of Korsgaard’s arguments of the CA, however—and I have significant doubts on that score—17—I think it’s clear that neither of these criticisms has much bite against Realism. I take them in turn.

Consider first the supposed problem of indeterminacy. As a criticism of the CA, we can see why it would be fatal if sound. The CA doesn’t just purport to prove the existence of a necessary being but of a particular one, God, with the particular attributes that God is supposed to have. To say in this context that the CA’s regress could just as well be terminated by Nature is indeed to subvert the CA’s intended conclusion. But even if the objection works in that case, the analogy does not carry over to Realism. The indeterminacy objection (whether targeting the CA or Realism) concedes that the regress has to end somewhere; it merely quarrels over where or with what. But Realism even as
Korsgaard construes it can easily survive such a quarrel, as long as we construe it as the name of a genus whose species constitute different answers about what ends the regress.

Granted, Korsgaard doesn’t conceptualize Realism this way; she associates it with an idiosyncratically Moorean (or Platonic) notion of “intrinsically normative entities,” and treats it as a view about precisely that sort of regress-stopper. But as we’ve seen, there’s no good reason to think of Realism in this way (or as exhausted by this conception), and even if we insisted on using the word “realism” this way, it wouldn’t change the basic issue. If Realism_K denotes the Platonic-Moorean view that regresses end with intrinsically normative entities (sensu Korsgaard), just think of Realism_K as one species of a larger genus of theories (call it whatever you want) which denotes the view that regresses end with regress-stoppers of some but any variety. Korsgaard’s indeterminacy objection is not an obstacle either to realism in the broad sense, or to foundationalist realisms in particular.

Now consider the charge that the Realist begs the question by assuming the very practical necessity he’s obliged to prove. The claim here is that to get a regress of moral justifications started, you have to assume that your starting point is in fact a moral requirement. So you begin ex hypothesi with “S must φ,” and then ask, “Why must S φ?” Korsgaard’s point is that the legitimacy of the justificatory “why?” question already presupposes that φ-ing needs justification, which in turn presupposes that it is the case that S must φ. The chain from regress to terminus then merely gives the appearance of justifying S’s φ-ing when in fact all it’s doing is drawing out the consequences implicit in the initial assumption that S must φ. So the procedure as a whole is circular (contrary
appearances notwithstanding). “S must φ” justifies the terminus, while the terminus justifies “S must φ.”

But it seems to me that there is a very simple way around this supposed difficulty, which is to reword the regress problem to eliminate the trouble-making circularity. Instead of beginning with “S must φ,” we can instead begin with “Prima facie, S must φ.” Instead of asking “Why must S φ?,” we can ask “Ex hypothesi, if it were true that S were required to φ, what would make that case?” Keep iterating the latter question, mutatis mutandis, until you reach a terminus. Whatever makes the terminus terminal will explain whether or not S must φ; there will be something about the terminus (if there is one) that makes further “why” questions impossible, unnecessary, or incoherent. “Prima facie, S must φ” gets the regress started but doesn’t by itself justify anything; it’s the terminus (whatever it is) that asymmetrically justifies “S must φ.” So there need not be circularity here. In fact, if anyone has begged the question, it is surely Korsgaard, not the realist: it is Korsgaard, after all, who asserts that realism “ends the regress by fiat” without discussing a single realist solution of the regress problem.

Indeed, the problem that Korsgaard faces seems to me worse than that. The objection she makes of the realist is that the realist fails to answer what I earlier called the prescriptively presuppositionless version of the normative question: is there anything that it’s necessary to do? But, as I argued in 7.1, it is Korsgaard who fails to answer that question. If so, Korsgaard’s criticism of Realism applies more obviously to her own view than to Realism.
3. Intrinsic normativity. This leads us more directly to difficulties in Korsgaard's definition of Realism itself. The second conjunct of Korsgaard's definition tells us that Realism is committed to "intrinsically normative entities," but it is quite unclear what this means. Korsgaard conveys its meaning in Sources principally by way of allusions to J. L. Mackie's so-called “argument from queerness” against (one form of) realism in his *Ethics: Inventing Right and Wrong*. But "queerness" was just Mackie's description of Plato's highly metaphorical account of the Form of the Good in the *Republic*, an account, we should remember, that arises in a passage of the *Republic* where Socrates, confessing ignorance as to the nature of the Good, paradoxically likens it to its "offspring, the Sun." Taken by itself, this is not much help in fixing the meaning of the concept.

Elsewhere, Korsgaard offers a more precise account of intrinsic normativity, based principally on Moore's (highly Platonic) conception of goodness in *Principia Ethica*. But the precision of this account of intrinsic normativity derives from its focus on Moore's theory, a focus not easily transferable to other realist theories.

This leads us to a dilemma for Korsgaard's account. Either intrinsic normativity is essential to her definition of realism or not. If it is essential, then given the obscurity of the notion, it needs to be filled out before we can even grasp what "realism" means. In this case, the more that Korsgaard relies (as she must) on some narrow, specialized interpretation of the "intrinsically normative entities" conjunct, the narrower the target of her critique of realism, and the less threatening to realism as such. The less she relies on a specific conception of realism, the less clear the conception of "intrinsic normativity" at issue, and the less clear her critique will be.
Suppose, however, that intrinsic normativity is not essential. In this case, were she to jettison the "intrinsic normativity" conjunct of her definition, she would be left merely with the claim that by realist strictures moral claims are true, full stop. In this case, her target would be what Geoffrey Sayre-McCord calls "minimal moral realism": (1) moral claims, taken literally, are literally true or false, and (2) some moral claims are true. This target, unlike the Moorean one, is sufficiently broad so that a successful attack on it would successfully put realism out the running as an answer to the normative question.

In fact, the text is indeterminate as between the Moorean and the minimalist interpretations of Korsgaard's critique of realism. Some of her remarks seem to require a commitment to the intrinsic normativity conjunct, and so narrow the scope of her attack, leaving other conceptions of unrealism untouched, hence undisputed, hence unexcluded as candidate answers to the normative question. Some of her remarks range more widely to suggest that her attack on realism is intended to target even the minimal form.

4. Procedural versus Substantive/Constructivism. As remarked above, having rejected realism, Korsgaard insists on distinguishing between PR and SR, affirming the former and rejecting the latter. But—and here I echo other commentators, such as Nagel, Cohon and Fitzpatrick—the distinction between SR and PR is so sketchy, ad hoc, and unclear that it is difficult to characterize or criticize.

PR is the view that there are answers to moral questions because there are correct procedures for arriving at those answers. SR is the view that there are answers to moral questions because there are moral facts or truths which those questions ask about; in other words, there are correct procedures for answering moral questions because there are moral truths or facts which exist independently of those procedures, and which the
procedures track. So the difference is one of causal directionality. On PR, moral truths are constituted by or constructed by procedures (as in Rawls’s constructivism); on SR, moral truths are tracked by procedures (as in Nozick’s truth-tracking epistemology, or similar reliabilist or virtue epistemic sorts of epistemology). Though I don’t find Rawlsian constructivism particularly plausible, this isn’t the place to rehearse the difficulties with it in any detail; a few brief points will have to suffice.

Korsgaard takes her own theory, and by implication Rawls’s, to endorse PR but reject SR. The reason for the rejection of SR is obvious, but the content and motivation of the endorsement of PR is unclear. Rawls says (slightly) more about moral truth than does Korsgaard, but all things considered, his view is highly obscure.

Korsgaard twice appeals to Rawls’s constructivism as a precursor of her own views on the relation between realism and constructivism, but given the obscurities of Rawls’s view, its iterations and variations over time, the differences in context between Korsgaard’s project and Rawls, and the differences between Korsgaard’s views and Rawls’s, the allusion is hardly clarifying.

Even if we adopt the broadest (i.e., most minimal) definition of moral realism (a la Sayre-McCord), it is difficult to see the intelligible sense in which Rawls’s or Korsgaard’s constructivism qualifies as “realist.” To appropriate the term “realism” to describe constructivism seems like appropriating “realism” to describe instrumentalism or idealism. At this point, the terminology has become too confused to be of any cognitive value.

Second, if on PR, procedures constitute moral truths, what criteria determine why procedure-set A is preferable to set B? A can’t be preferable to B on grounds of truth-
conducivity (or anything parasitic on or related to it), but truth-conducivity is not the criterion. Korsgaard leaves it unclear what else will do the trick.

Finally, notice that while SR takes procedures to be certified as “correct” by their truth-conducivity, no defender of SR would take even the best procedures to be infallible in that respect. SR procedures are such that, if a well-functioning agent were diligently to follow them in optimal or favorable circumstances, it will probabilistically tend to be the case that the agent tracks moral truth. This is of course a tall epistemic order, and we can haggle about precisely what procedures and/or circumstances will do the job, and with what probability of success. What is particularly implausible about PR is that while PR procedures don’t track truth, they are infallibly correct in some other, non-truth-conducive sense. What is obscure is what this sense is and how infallibility is achieved with respect to it.

5. Realism and regresses. One of the curious features of Korsgaard's critique of realism is its relation to what I described in Chapter 1 as the meta-ethical regress of justification. Though she initially defines realism as a response to voluntarism, Korsgaard's criticisms suggest that realism is at a fundamental level a sort of abortive response to—a failure sufficiently to grapple with—the justificatory regress problem. The fundamental problem with realism is its supposed failure to acknowledge the reality or urgency of the regress problem and a fortiori a failure to identify the terminus of the regress.

The oddity here is that this criticism is very difficult to make coherent with the rest of what Korsgaard has to say about normativity in Sources. If realism fails for its failure to address the regress problem, that implies that the regress problem is the
fundamental formulation of the normative question. But while Korsgaard discusses the regress problem in *Sources* and elsewhere, nothing about her discussion of the normative question suggests that she takes the content of the question fundamentally to be determined by that problem. It plays no significant role in Lecture 1’s discussion of the question, and no significant role anywhere in *Sources* as being the agenda-setting issue in ethics. There is in fact no index entry devoted to it, either. A puzzle therefore presents itself: why is the regress problem the fatal problem for realism but not the fundamental issue to which Korsgaard’s theory is intended to resolve?

Further, if realism fails because it has no answer to the regress problem, a theory succeeds, presumably, if it has one. It is not clear whether Korsgaard takes her own theory implicitly to have one, but either way, we face an insuperable problem. If she has no answer to the regress problem, her theory falls prey to precisely the same criticism as befell realism. Let’s suppose then that her answer to the normative question is intended as an answer to the regress problem. Apart from the fact that it isn’t at all discussed in this way, we face a new problem: now, the distinction between SR and PR in this light becomes utterly incomprehensible even apart from the more abstract criticisms I leveled against it above. SR is the view that there are answers to moral questions because there are moral facts or truths which those questions ask about. Let us say *ex hypothesi* that Korsgaard’s theory answers the regress problem with some answer, A. In that case, it is trivially true that A is the answer to the normative question because A terminates a regress of practical justification. But it must do so in virtue of something about it—some facts—which make it the case that A (as opposed to something else, e.g., the “intrinsically normative entities” of the realist) is capable of terminating the regress in the
right way. Surely then to the extent that Korsgaard's answer is the answer because it terminates the regress, the answer is true because there are (regress-related) facts that make it true. If there were no such facts, there would be nothing to make Korsgaard's answer more terminus-worthy than, say, Aristotle's, Aquinas's, Mill's, or Moore's. But then Korsgaard's theory is itself an instance of SR, and the SR/PR distinction is yet further undermined, as is Korsgaard's critique of realism. 20

7.5 Notes

1 A typical example is Michael Smith, The Moral Problem (Malden, MA: Blackwell, 1994), which begins by discussing what Smith plainly regards as "the" uniquely fundamental problem of ethics, only to settle the status of the problem by stipulation (p. 11).


3 Questions (2), (4), and (5) could in principle be given weaker readings in terms of reasons for action.

    (2*) Is there anything that beings like me have reason to do?
    (4*) Why is it that I have reason to ~X when desire bids me to do X?
    (5*) When desire bids do I have reason to follow its promptings or not?

I ignore these readings, however. For the weaker our reading of the question, the weaker Korsgaard’s answer turns out to be, and the weaker it is, the less dialectical bite it has as a competitor to the meta-ethical egoism I’ve defended elsewhere in this study.

4 Question (1) makes reference to moral concepts, not morality.

5 In his commentary “History, Morality, and the Test of Reflection,” in Korsgaard, Sources of Normativity, pp. 210-11, Bernard Williams counts two questions, which he calls (N1), (N2). (He counts a third question, [R], but neither Williams nor I
take this to be an instance of the normative question, proper.) His (N1) corresponds to my (6) and his (N2) corresponds to my (2). Two points are of interest here.

First, it’s telling that Williams sees two questions where I see seven; if I’m right to distinguish the seven questions, that suggests the ease with which even careful readers might conflate Korsgaard’s versions of the normative question.

Second, it is interesting to note that while (the wording of) Williams’s questions correspond to mine, his interpretations of the questions and of the differences between them are significantly different. “[N2] is at least broader than [N1],” he writes, “since there are non-moral forms of normativity. But this does not seem to allow enough for non-moral forms of normativity (prudential, aesthetic, etc.) which, like the moral sort, can equally give trouble with inclination. It is not entirely clear to me whether Korsgaard thinks that there is a problem about the nature of normativity before we ever get to the specifics of morality.”

Two further points. First, what Williams takes to be a matter of “breadth” I have argued in the text to be a matter of fundamentality. It is not merely that (N1) happens to be broader than (N2), but that the answer to (N1) presupposes the answer to (N2). Second, unlike Williams, I do not take it to be an implication of (N2)’s greater breadth that (N2) implies a commitment to “non-moral forms of normativity.” (N2) simply asks whether there are any actions that, all things considered, we must do. The question is intelligible whether moral and non-moral reasons are distinct and incommensurable, or whether moral reasons end up being overriding (hence what, all things considered, we must do), or finally, whether morality is just a name for “those reasons which, all things considered, we must do.” Contrary to Williams, nothing about the sheer formulation of the question settles this.


7 The locus classicus is Socrates’ denial of the possibility of akrasia as expressed in Plato’s Protagoras, 352a-362a, and to some extent Plato, Republic, VI, which by implication amounts to the claim that moral knowledge is sufficient for virtue. Moral knowledge here, however, is knowledge (not of propositions but) of Forms, so that S’s knowledge of (the Form of the) Good is sufficient for being motivated to seek it. Now, Platonic Forms are of course not “concepts,” but being more analogous to concepts than judgments, they are a closer approximation to the claim Korsgaard would need to bridge the gap between the moral concept and normative force questions. See Terence Irwin, Plato’s Ethics (Oxford: Oxford University Press, 1995), secs. 58 and 165.

8 Korsgaard, Sources of Normativity, pp. 42-43; the discussion of Prichard goes from pp. 38-42.


Korsgaard, *Sources of Normativity*, pp. 78-86.

Cf. Korsgaard, “Skepticism about Practical Reason,” in Korsgaard, *Creating the Kingdom of Ends*, p. 321, where she herself seems to recognize this point.


To put the point more precisely but more awkwardly, Korsgaard’s claim is ambiguous between the following four interpretations. Each of us:

(a) has a practical identity
(b) has and sustains a practical identity
(c) has and ought to sustain a practical identity
(d) ought to come to have and ought to sustain a practical identity.

Scholarly commentary on *Sources of Normativity* involves a discrepant and incongruous set of attitudes on this subject. Many commentators simply ignore the question of the cogency of Korsgaard's taxonomy altogether, taking it for granted despite its unorthodoxy and counter-intuitiveness and despite its centrality to her argumentative strategy. Oddly, some of the same commentators then remark on the "frustrating" and "obscure" qualities of the book (e.g., Gibbard, Donner) without remarking on what would appear to be its most obvious source. Donner's review is particularly incongruous in this regard, praising Korsgaard for the "clarity and elegance" (Wendy Donner, Review of *The Sources of Normativity*, *Dialogue*, vol. 36, no. 3 [Summer 1999], p. 653) of her classificatory scheme, but then taking strong issue with Korsgaard's interpretations of Hume and Mill (p. 655). But part of the difficulty with Korsgaard's treatment of those figures is precisely how she categorizes them!

In pattern, most commentators take the taxonomy for granted but dispute the treatment of a specific theory or figure within it. Thus Nagel, Cohn, and Fitzpatrick each take issue with Korsgaard's treatment of realism and Nagel takes issue as well with her treatment of ethical egoism. Williams takes issue with her account of Hume and of his own theory. Cohen, Geuss, and others take issue with her interpretation of Kant. Though I am not aware of any scholarly commentary on *Sources of Normativity* by specialists in ancient philosophy, Korsgaard’s interpretations of Plato and Aristotle are obviously
controversial. But interesting as they are, these disputes are of little dialectical interest except insofar as they contribute to or detract from Korsgaard's overall argument. And yet, to my knowledge, no commentator has taken issue with her taxonomy as such, surely the more dialectically fundamental issue.


17 Korsgaard’s account of the CA comes from the very cursory discussion devoted to it in Part IX of *Hume’s Dialogues Concerning Natural Religion*. For a more sophisticated version of the argument, see William Lane Craig, *The Kalam Cosmological Argument* (Eugene, OR: Wipf and Stock Publishers, 1979). Though I don’t think Craig’s argument is sound construed as an argument for the existence of God, the part of his argument that establishes the finitude of past time is strikingly analogous to the best arguments in defense of epistemic and meta-ethical foundationalism. Korsgaard’s arguments touch neither Craig’s argument for the CA nor the standard (much less the best) arguments for foundationalism.


19 Korsgaard, *Sources of Normativity*, pp. 19 and 35.

20 Thanks to an audience at Bowling Green State University (March 2004), Michael Young, Jason Raibley, and Carrie-Ann Biondi for helpful feedback.
8.1 Introduction

In the previous chapter I took issue with Korsgaard's formulation of the normative question. In this chapter, I take issue with the first part of her answer, namely, that reflection furnishes the source of normativity. It might help, then, to begin with an overview of this latter part of her argument.

We begin, for her, with three bedrock facts about human nature, assumed but not argued for in Sources: (i) first, that human beings have first-order desires, in Frankfurt's sense of that term; (ii) that we have a capacity for self-consciousness that takes first-order desires as objects; and (iii) that having activated our capacity for self-consciousness, we have the capacity to ascend (again, in Frankfurtian terms) to higher-order desires on our first-order desires, i.e., to form desires about desires. According to Korsgaard, in ascending to an n-order desire higher than the first-order, the agent creates motivational “distance”² between his first- and n-order desires, giving rise to a "normative problem" requiring solution.

In creating this distance, the agent's ascent to the n-order level puts her first-order desires in question, raising for the first time whether or not the agent ought to act on those desires. The agent then either endorses the desire or not, where the act of endorsement is
the act of identifying oneself (one's practical identity) with the principle that the desire expresses. This very general commitment to principle then comes to govern the agent; since it arises autonomously, it leaves the agent's freedom in place. Since it genuinely governs the agent and constrains her actions, it is normative. Korsgaard thus concludes that self-consciousness—or to use her preferred term, "reflection"—is the source of normativity, the capacity whose exercise both gives rise to and resolves the normative problem.

Despite the ingenuity of this argument, I do not think it succeeds. In fact, I think it fails at virtually every step. To show this, I want to proceed slowly through the argument, beginning with its Frankfurtian presuppositions and moving in a detailed way through each stage of the argument to its conclusion. I want to suggest that it's precisely the Frankfurtian component of the argument that is at once unique and problematic: in some respects, Korsgaard unwittingly inherits problems from Frankfurt's account that beset her own; in other respects, she borrows features of Frankfurt's account that give the appearance of bolstering her argument, but in fact don't.

8.2 Korsgaardian Meta-Ethics, Frankfurtian Psychology: Some Preliminaries

Since Korsgaard's reliance on Frankfurt's work is, as she puts it, "obvious," it will be useful to have an overview of as much of Frankfurt's theory as is relevant to an interpretation of Korsgaard's. The central text is Frankfurt's "Freedom of the Will and the Concept of a Person," of which two elements are essential. The first is Frankfurt's taxonomy of desires, which consists of a distinction between types of desires as well as an account of the hierarchies that obtain among desires. The second is an account of the
bearers of desire, that is, an account of the various entities capable of generating different sorts of desiderative hierarchies. I take these in turn.

*Frankfurt's taxonomy of desires.* What sorts of desires are there? A first-order desire, according to Frankfurt, is any desire captured by a statement of the form, "A wants to X," where A refers to an agent and X refers to an action. Frankfurt notes early on that the concept of "wanting" or desiring is "extraordinarily elusive," but a basic distinction within the set of first-order desires is that between effective and non-effective desires. An *effective* desire is a desire the possession of which does, will, or would move its possessor all the way to action; a *non-effective* desire inclines the agent toward a course of action without moving him all the way there. Frankfurt's basic unit of analysis is a first-order effective desire, which he identifies with the possessor's will. Having noted this, we can from now on set non-effective desires to one side and focus on effective desires, or wills.

A second-order desire is a desire that takes a first-order desire as its object—doing so by analogy with the way in which a first-order desire takes an action as its object. So second-order desires are described by statements of the form, "A wants to X," where X is a first-order desire; alternatively, we might say that a second-order desire is one described by a statement of the form, "A wants to want to X," where X is an action. Here, too, we can distinguish between effective and non-effective desires. An effective second-order desire is one in which an agent has a certain first-order desire and, from the second-order level, wants that desire to be his will, i.e., wants it to be effective. "It is not merely that he wants the desire to X to be among the desires by which, to one degree or another, he is moved or inclined to act. He wants this desire to be effective—that is, to
provide the motive in what he actually does.”7 In Frankfurt's terminology, an effective second-order desire is (by contrast with a non-effective one), a second-order volition.

The bearers of desire. What sort of entity has what sort of desire? The basic distinction, on Frankfurt's account, is that between wantons and persons.

A wanton is an entity capable of having first-order desires (both effective and non-effective) but without ever having second-order volitions. A wanton, in other words, has a set of first-order desires, and at any given time, acts on whatever desire at a given time happens to be strongest, i.e., effective. Wantons either lack the capacity to generate—or simply do not generate—second-order volitions on their first-order desires.

A person, by contrast, is an entity possessing both first-order desires and n-order desires above that. A person so conceived is voluntarily capable of ascending to n-order levels on its first-order desires, and of governing her first-order desires from the higher-order set. The distinctive attribute of persons, then, absent in wantons, is the capacity for freedom: a person is free, on Frankfurt’s view, when her n-order volitions coincide with her first-order desires, and unfree when they don’t.

With this primer account of Frankfurt's theory in hand, we can now consider Korsgaard's use of it. As will soon become apparent, the primer account of Frankfurt will have to be expanded and modified as we get farther into the complexities of Korsgaard's argument.

8.3 Problem (1): The Normativity of First-Order Desires

Korsgaard’s account, as we’ve seen, begins with what I’ve called the bedrock fact that human beings have first-order desires and a capacity for reflection on them. The
theory takes that assumption as its starting point, and proceeds to account for normativity by appealing to facts that arise as a consequence of our possession of (effective) first-order desires. Korsgaard thus takes for granted that we have such desires and act on them, that our doing so is an unproblematic fact requiring no further elaboration, and that the source of normativity is to be identified by taking first-order desires as given, and of focusing on what we do in consequence of having them. The implication is that Korsgaard's proffered source of normativity will not account for—is not intended to account for—the normativity of first-order desires. Their normativity is assumed to be unproblematic, hence unrelated to what she takes to be the normative problem she intends to solve.

I want to suggest that this seemingly innocuous assumption is a mistake—in fact one mistake with three distinct aspects. In taking the genesis of first-order desire for granted, Korsgaard (a) leaves the fundamental version of the normative question (discussed in Chapter 7) unanswered; (b) leaves the normativity of first-order desires unexplained, and (c) jeopardizes the desired unity of her account of normativity.

To see this, consider the following dilemma generated by Korsgaard’s account. The point of her theory is to account for normativity which, as I argued in Chapter 7, is an account of why we’re required to act at all. Korsgaard’s theory begins, however, with the fact that we have first-order desires, which, qua effective desires, move us to act and causally explain why we do act. Now, either (1) first-order desires have some sort of normativity (i.e., give reason for action), however prima facie or attenuated, or (2) they have none.
Suppose that (1) is the case. In this case, the theory’s starting point is *ipso facto* in tension with its purpose. The point of the theory is to explain normativity, but its starting point, in assuming the normativity of first-order desires, precludes an account of that very thing.

Suppose then that (2) is the case. In this case, the agent begins with a set of first-order desires which, qua desires, have no normative force whatsoever. On (2), S’s merely *having D* gives S no reason to satisfy D. D is, as it were, a normatively neutral fact about S; it’s just “there.” Desires so construed may *impel* action, but don’t constitute *reasons* for action. Korsgaard then asserts that S comes to reflect on D, perhaps to form a second-order desire that takes D as its object, running D through a reflective test to determine whether or not D is to be satisfied. If D passes the test, D is normative (and thus can be satisfied), whereas if D fails the test, it isn’t (and can’t).

The difficulty is this: how can a first-order desires that, *ab initio*, has no normativity come to acquire normativity from a second-order operation on the desire, when the content of the second-order operation is determined by and parasitic on the first-order desire itself? Recall that reflection is simply a higher-order mental act that takes a first-order desire as its object. So the content of reflection is simply the desire itself, seen from a certain perspective or distance. How does a desire that had no normative force come to acquire it merely because it passes a test of reflective endorsement? What *feature* of the desire is being endorsed? Suppose that I want to eat a sandwich at t₁, and suppose *ex hypothesi* that the fact (at t₁) that I want to eat it doesn’t give me a reason to eat it. How then is reflecting on my desire to eat it going to give me a reason to eat it? Reflection may uncover facts—about the would-be sandwich, about me,
about my situation—that were opaque to me at the first-order level. But they cannot create normativity *ex nihilo*.  

This difficulty suggests an ambiguity in Korsgaard’s Frankfurtian psychology, one widely noted by commentators and (as we’ll see) of some relevance in interpreting Korsgaard. The issue is the precise role of belief versus desire in Frankfurt’s theory, and commentators characteristically distinguish between conative and cognitive interpretations of the theory as regards the precise nature of the ascent from first- to second-order desires. Frankfurt’s paper is systematically ambiguous as between these two interpretations.

On the *cognitive* interpretation, second-order desires arise when the agent forms the appropriate evaluative belief about his first-order desires. So I might have a first-order desire, believe it to be worth satisfying, and thereby form a second-order desire as a consequence of the formation of that evaluative belief. The order of ascent is: desire, belief, desire. The belief mediates between first- and second-order desires.

On the *conative* interpretation, second-order desires arise immediately from first-order desires without the mediation of evaluative beliefs. On this interpretation, I have a first-order desire, and without having to form an evaluative belief, I come to form a second-order desire that takes the first-order desire as its object. Here the ascent is directly from desire to desire.

The cognitive/conative distinction gives us two possibilities for conceptualizing the relation between reflection and normativity.

*Cognitive interpretation*: The act of reflection involves the formation of an evaluative belief about the agent’s first-order desire, where this belief, in turn, guides the formation of a second-order volition.
Conative interpretation: The act of reflection immediately consists in the formation of a second-order volition without the mediating link of an evaluative belief.

Whichever possibility we pick leads us to a dilemma; either possibility really gives rise to twin variations on the same dilemma.

Take the cognitivist interpretation. In this case, \( S \) has \( D \), and ascends to reflection on \( D \), which means that he forms a higher-order evaluative belief \( B_e \) that takes \( D \) as its object. But plausibly, \( B_e \)'s taking \( D \) as its object means that \( B_e \) consists of an evaluation or assessment of the desirability of \( D \)'s being satisfied. If \( S \) judges \( D \) to be worth satisfying, \( S \) can legitimately satisfy \( D \), and if not, not. But for \( S \) to judge \( D \) to be worth satisfying, there has to be something about the satisfaction of \( D \) that makes it worth satisfying. The something may be opaque to \( S \) until and unless he reflects on it, but while what is opaque to him may be epistemically unavailable at the first-order level, the “something” cannot literally be conferred by the fact of reflection or evaluative belief-formation. A judgment about the desirability of satisfying a desire is a judgment about the features of the situation in virtue of which the desire is worth satisfying. If the features aren’t there, it’s not as though an evaluative belief can put them there, and if the features are there, it’s not as though an evaluative belief can take them away. For this reason, Korsgaard’s failure to account for what it is about features of situations that agents find themselves in that makes first-order desires normative for them leaves an air of mystery in her account. If the italicized item is unaccounted-for, it’s unclear how evaluative beliefs, simply as second-order operations, will account for their normativity.
The conative interpretation leads to more or less the same problem from a slightly different direction. On this interpretation, the agent forms a second-order desire in relation to his first-order desire without the intervening step of the higher-order evaluative belief. The second-order desire is on this view simply an immediate consequence of reflection, and one that doesn’t require an intervening belief for its formation. For this interpretation to be intelligible, the second-order desire has to be formed as a response to the desirability or undesirability of the first-order desire. In other words, the mechanism of desire-formation non-cognitively does the work here that evaluative beliefs do on the cognitivist interpretation.

If so, we simply face an analogue of the problem here as we saw in the cognitivist case. The second-order volition has to be a response to the desirability (or not) of the first-order desire, and if so, it has to respond to something antecedent to and unconferrered by the mechanism of desire-formation. In other words, second-order volitions respond, non-cognitively, to the desirability of what’s desired. But in taking the normativity of first-order desires for granted, that is precisely what Korsgaard has left unexplained. If first-order desires qua first-order have no particular normative status, it is unclear what conative second-order volitions are responding to in order to produce second-order desires about them.

We can amplify the force of the preceding problem if we note just how much is left unaccounted-for by the failure to account for the normativity of first-order desires. I noted above that Frankfurt offers a “taxonomy” of desires, but it’s important not to think of this taxonomy as either being or purporting to be exhaustive of the major types of desire. Frankfurt himself notes that his taxonomy falls far short of this:
People are generally more complicated than my sketchy account of the structure of a person’s will may suggest. There is as much opportunity for ambivalence, conflict and self-deception with regard to desires of the second order, for example, as there is with regard to first-order desires.\textsuperscript{11}

Frankfurt goes on at this point to discuss the complexities of second-order desires, but what is easy to overlook (because merely mentioned in passing) is the implicit complexity of first-order desires.

It is tempting, given the grammar of Frankfurt’s formula for first-order desires—“I want to $X$”—to think of the value for $X$ as something very simple: “I want to eat,” “I want to sleep,” “I want sex,” etc. But nothing about the nature of first-order desires requires this. First-order desires can be as complex as imaginable; the only form of complexity excluded to them is their (a) taking other desires as objects, and (b) taking objects that require the taking of other desires as objects (e.g., desiring an object the very desiring of which leads naturally to second-order desires). It’s far from obvious what the conjunction of (a) and (b) amounts to, and how it constrains the complexity of first-order desires, but it gives us no reason to think them simple.

In taking the normativity of first-order desires for granted, then, Korsgaard is taking for granted the normativity of three items:

(1) Simple first-order desires of the form, “I want to eat,” “I want to sleep,” etc.
(2) Complex first-order desires of the form, “For an appetizer, I would like the oysters mignonette, followed by the grilled filet of beef with gnocci’s—make that medium rare—with a red wine, the Chambolle Musigny Ler Cru, ‘Les Carrieres’; and then we’ll need to see the dessert menu.”
(3) The complex causal or teleological relationships between simple and complex first-order desires: “After that meal, I’ll have to take a nap,” or “In order to afford this meal, we’re going to have to break the bank.”
These omissions are not innocuous, and they raise three distinct but interrelated problems for Korsgaard’s theory. At one level, they simply leave a lot unexplained, and in so doing, leave Korsgaard’s original (“fundamental”) normative question unanswered. At a somewhat deeper level, they threaten the supposed unity of her account. But at an even more fundamental level that subsumes the preceding two, they raise the question of whether, in locating the source of normativity in a second-order operation on first-order desires, Korsgaard can really account for the source of normativity at all.

Plausibly, one source for the normativity of first-order desires is need-satisfaction, understood in the context of an Aristotelian conception of flourishing or well-being. If so, first-order desires would exist for the sake of the contribution they made to the agent’s flourishing: flourishing explains why the agent needs to form first-order desires in the first place. Reflection might play an epistemic role in uncovering facts about desires—for instance, whether certain desires (or more precisely, the formation of certain desires vis-à-vis objects of pursuit) were appropriate to the agent’s flourishing. Reflection could in fact be a constituent of the agent’s flourishing itself, so that reflection not only uncovered facts about the desirability of an object, but also enhanced its desirability. But reflection wouldn’t literally constitute facts about desirability any more than it constituted facts about needs, flourishing, or well-being. Need-satisfaction is a causal relationship between an agent’s flourishing and the objects she pursues; whatever role reflection played in need-satisfaction, it wouldn’t literally produce the truth-conditions of such relationships.

The Aristotelian account is not, of course, the only possible account of the normativity of first-order desires. But *ex hypothesi*, suppose that we accept it, or
something like it. In that case, Aristotelian flourishing would explain the normativity of first-order desires: it would explain what first-order desires were for, their *raison d'être*.

I’ve suggested above that Korsgaard leaves the normativity of first-order desires unaccounted for. Notice, however, that were she to avail herself of, say, the Aristotelian account, her doing so would produce a problematic bifurcation in her theory as a whole. On the one hand, the Aristotelian account locates normativity in the teleological character of human flourishing. On the other hand, Korsgaard’s officially Kantian account locates the source of normativity in the act of reflection. The two theories would invariably compete with one another and introduce an incompatibility in Korsgaard’s account of normativity.

In fact, I think this is precisely what happens. Lecture 3 locates the source of normativity in reflection, citing a specifically Kantian inspiration for the claim. But in the course of a response to an objection she entertains in Lecture 4—an objection where the normativity of first-order desires is at issue—Korsgaard shifts ground and adopts just the sort of Aristotelian conception of flourishing I’ve described in the foregoing.

The best account of what an animal is comes from Aristotle. We have already seen that Aristotle thought that the form of a thing is the organization or arrangement of its parts that allows it to be what it is to do what it does, to do its job. Now Aristotle thought that a *living* thing is a thing with a special kind of form. A living thing is so designed as to maintain and reproduce itself. It has what we might call a self-maintaining form. So it is its own end; its job is just to keep on being what it is. Its business in life is to preserve its own *identity*. And its organs and activities are arranged to that end.

Here, the source of normativity is no longer reflection, but Aristotelian flourishing, or survival qua human. First-order desires exist *for the sake of* teleological self-maintenance.
Korsgaard has three options here, none of them palatable. She could leave the normativity of first-order desires unexplained, but if so, it remains a mystery how second-order volitions could be the source of normativity if they depend for their content on first-order desires. She could produce some Kantian account for them, but she doesn’t. Or she could explain the normativity of first-order desires by way of an Aristotelian-type account, as in fact she ends up doing.

But if Korsgaard takes the Aristotelian option—and wants a unified account of the source of normativity (as she says she does)—she cannot say that reflection is the source of normativity. On this option, Aristotelian flourishing is. On the other hand, if she wants to insist on reflection’s being the source of normativity, flourishing can’t be; reflection is. The result is a serious incoherence in her theory. Korsgaard is committed to saying that Kantian reflection and Aristotelian flourishing are each the source of normativity, despite their incompatibility; and in her Kantian moments, she is committed to saying that reflection, which depends for its very operation on first-order desires, is the source of normativity in the absence of any account of the normativity of first-order desires. Neither claim is tenable, taken separately, and each is inconsistent with the other.

8.4 Problem (2): Why Reflect?

Suppose that we set the criticism of section 8.2 aside and grant Korsgaard’s assumption that possession of first-order desires is bedrock for normative theory. The second of Korsgaard’s bedrock claims is plausible enough: that we have a capacity for reflection. According to Korsgaard, the act of reflection is the paradigmatic human act of freedom. It is therefore up to us, and is also the source of normativity, explaining and
giving rise to moral requirements. The difficulty is this: if the act of reflecting—of actualizing our capacity for reflection—is up to us, the agent need not engage in it. If the agent need not engage in it, however, in what sense is it a requirement? Why must we reflect?

Before addressing this question directly, we need to consider a metaphysical complication concerning Korsgaard’s view of freedom. Korsgaard regards reflection as our paradigmatically free act; indeed, she offers two accounts of normativity, one invoking reflection, and another equivalent one invoking freedom. But her (Kantian) account of freedom is compatibilist, and so is consistent with determinism.17 A libertarian might therefore be tempted to pick a metaphysical quarrel with Korsgaard on that issue. Though I am sympathetic to that criticism, for present purposes, I propose to sidestep the compatibilist/libertarian dispute and take Korsgaard’s claims about reflection and freedom at face value. To say that reflection is “free” is to say that the act of reflection is up to the agent, or in the agent’s power, however the italicized terms are to be understood. So reflection is an act, not a mere happening; it is something that the agent does, not merely an event she undergoes.

This leads us directly to the difficulty. If reflection is an act, and the point of an account of normativity is to justify why we’re required to act, what requires that we reflect? Insofar as Korsgaard deals with it,18 her account seems to involve a problematic slide from capacities to necessities. The argument begins with

(1) S has the capacity to reflect.

From (1), it moves to

(2) S does reflect.
And from (2) to (3):

(3) S’s reflecting is the source of normativity (for S).

Korsgaard clearly takes (3) to entail (4):

(4) S’s reflecting is the basis of the moral obligations that bind S,
which is equivalent to saying that reflection is the basis of moral requirements.

It’s not clear whether Korsgaard would endorse (5):

(5) S is required to reflect.

But if she doesn’t endorse (5), we’re left with the following question: if S is under no obligation to reflect, and has the power not to reflect, then S need not reflect—thus jeopardizing the move from (1) to (2). If that move is put in jeopardy, so are its successors, and if that is the case, we reach the conclusion that S is not bound by moral obligations. It follows that S’s failure to reflect is a conceivable way of S’s escaping moral obligation: after all, S’s reflection being a free act entails that S can escape the consequences of reflection by not reflecting. This result is surely incompatible with the categorical conception of moral obligation that Korsgaard espouses, and even as a hypothetical or conditional conception, it leaves a great deal unexplained.

My main concern is with the fallacious (or at least unmotivated) move from (1) to (2). We see it in passages like the following:

Normative concepts exist because human beings have normative problems. And we have normative problems because we are self-conscious rational animals capable of reflecting about what we ought to believe and do. That is why the normative question can be raised in the first place: because even when we are inclined to believe that something is right and to some extent feel ourselves moved to do it we can still always ask: but is this really true? And must I really do this?
I’ve italicized the words that indicate the shift. In the first instance, Korsgaard tells us that we are capable of self-consciousness. But the possession of a capacity does not explain why its possessor should, ought, or must actualize it. In the second instance, it’s true that we can ask the questions she mentions, but it is important to see how much work is being done for Korsgaard’s account by the seemingly innocuous concession that we do in fact ask them. Asking the question is what leads to the normative problem; the existence of the problem calls for the need for a solution and thereby shows (by Korsgaard’s lights) that the capacity for asking such questions is “the source of normativity.” But we can block the whole line of argument by asking why the agent is required to rise above his first-order desires to reflect on them and ask such questions in the first place. If the agent resolutely stays a wanton (not by reflecting on the desirability of staying one, but simply by doing so), he faces no normative problem, is bound by no part of the solution—and a fortiori is bound by no part of morality.

The problem becomes clearer in a somewhat longer and more explicit passage in Lecture 3:

And this [our capacity for self-consciousness] sets us a problem no other animal has. It is the problem of the normative. For our capacity to turn our attention on to our mental activities is also a capacity to distance ourselves from them, and to call them into question. I perceive, and I find myself with a powerful impulse to believe. 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 believe? Is this perception really a reason to believe? I desire 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, not just as such. It needs a reason. Otherwise, at least as long as it reflects, it cannot commit itself or go forward. If the problem springs from reflection then the solution must do so as well. If the
problem is that our perceptions and desires might not withstand scrutiny, then the solution is that they might. We need reasons because our impulses must be able to withstand reflective scrutiny. We have reasons if they do.\textsuperscript{21} This passage begins with claims about what we can do but ends with claims about what we must do. But the connection between “can” and “must” is supplied by nothing more than the tacit assumption that if we can reflect, we must reflect, and once we do, we’re locked into moral obligation.

A capacity to turn attention on to my mental activities—or to distance myself from them—gives me no reason to do so. Even if we grant that the reflective mind cannot settle for perception and desire, Korsgaard has not addressed the fact that an unreflective mind can, and if it can, it has no reason to become reflective. If “backing up” is what causes the normative problem to arise, not backing up will keep it from arising. If perception and desire do not “withstand scrutiny” from the reflective standpoint, it follows that there is no scrutiny to withstand if one never ascends to that standpoint.

So, contrary to appearances, we’ve been given no reason to take the first step toward generating the normative problem to which Korsgaard’s theory is the solution. And without justifying the initial step to reflection, Korsgaard has left the most crucial step of her argument undefended. We must either assume that the agent does reflect or must reflect to get the theory off the ground. As I’ve suggested, Korsgaard clearly does not show that the agent must reflect, and in identifying reflection with freedom, she is debarred from assuming that the agent is somehow driven to, or merely brutely comes to, reflect in the way that (we are assuming \textit{ex hypothesi} in this section) that the agent is driven to acquire or brutely comes to acquire first-order desires. Precisely because
reflection is up to the agent, we need a justificatory, not merely an explanatory, story to account for it.

Recall that the story, in compliance with Korsgaard’s own “address” requirement, has to appeal to the agent’s own perspective. But *ex hypothesi*, the relevant agent is an unreflective wanton. And it surely makes no sense to say that a wanton does reflect because he has the capacity to do so; to say that is to say that a wanton ceases to be a wanton because she has the capacity to cease being one. What makes a wanton a wanton is what she does, not what she has the capacity to do. So if we grant that adult wantons are possible (as Korsgaard does), Korsgaard’s account of normativity fails her own address requirement with respect to them.

Korsgaard might conceivably let go of that assumption, and claim instead that adult wantonhood is nomologically (i.e., psychologically) impossible. But to claim that is really just to say that human adults are, at a certain point, causally pushed into reflection, i.e., that we just come to reflect and can’t help but do so. That’s possible, but in that case, reflection cannot be freely up to the agent, as Korsgaard insists it is. And Korsgaard’s giving up that assumption would be tantamount to her giving up the theory as such.

8.5 Problem (3): *What Normative Problem?*

In Section 8.4, I argued that Korsgaard had given us no reason to think that reflection is a practical requirement. But let’s set this criticism aside and assume that the agent *does* reflect. As we’ve seen from the passages quoted above, Korsgaard now claims that the agent’s reflecting gives rise to motivational distance, which in turn gives rise to a “normative problem” solved by the categorical imperative. In this section, I want to
dispute the idea that reflection necessarily gives rise to any such problem (or indeed, to any problem at all). Since it need not do so (I argue), we need not be inexorably led by it to the categorical imperative in the way that Korsgaard suggests we are.

We begin for Korsgaard with the assumption that the agent has a repertoire of first-order desires. The agent then comes to reflect on these desires. As Korsgaard (implicitly following Frankfurt) describes the phenomenology, reflection weakens the agent’s attachment to his first-order desires, introduces “distance” between his agency and the desire, diminishes the desire’s “domination” of his agency, and calls the normative status of the desire into question. Having done so, the agent then infers that he faces a problem, “the normative problem.” At $t_1$, he had a first-order desire that he didn’t question—hence faced no problem. At $t_2$, he reflects. At $t_3$, on reflection, he notices that he need not act on the first-order desire. And now, Korsgaard infers, he asks the question, “Shall I act on the desire?”—thereby facing the problem of how to answer that question and, generally, how to act. An exercise of reflection produces the answer.

The point to notice about this account is that its plausibility derives almost exclusively from the peculiarities of Korsgaard’s description of the phenomenology of reflection. Were we to describe that phenomenology differently (as I’ll do below), we could very easily eliminate the “problem” that Korsgaard believes she has discovered—and in so doing, eliminate the need for the solution she invokes.

First, Korsgaard consistently describes first-order desires as alien “impulses” that “dominate” the agent. This follows Kant’s conception of the material principles of the will in the *Critique of Practical Reason*, which conceives of our desires as “mere” inclinations to act, blind hedonistic impulses that grip and manipulate us by analogy with
the way in which we might be manipulated by the forces of nature or by other persons. Given this description, we are, I think, covertly led into the supposition that a wanton (i.e., a possessor of nothing-but first-order desires) is in an adverse situation purely by virtue of being a wanton. Indeed, the alarming picture painted is that the wanton is effectively bullied by his desires. This picture creates the *prima facie* presumption that a wanton’s situation is somehow untenable as such, and that he *needs* to do something to get himself out of it.\(^{28}\)

But it’s important to see that whatever impression one gets of practical necessity from this picture is an artifact of Korsgaard’s description of the phenomenology, as inherited from Frankfurt’s description of the same phenomenology. We’ve been given no argument whatsoever for thinking of desires as slavemasters, and no argument for thinking that it’s bad to be their slave. And so, we’ve been given no argument as yet for thinking that wantonhood constitutes a problem requiring solution.\(^{29}\)

In my view, this difficulty arises from unclarities in the account of wantonhood that Korsgaard inherits from Frankfurt. Frankfurt is radically unclear about what entities in the world actually qualify as being wantons, and for the same reason he is unclear about the cognitive powers wantons are supposed to have. In adopting Frankfurt’s account, Korsgaard inherits the unclarities as well.

A wanton, we learn from Frankfurt, is an entity with first-order desires but no second-order volitions. This formulation prompts two interrelated questions. First, can wantons have (non-volitional) second-order desires? Some of Frankfurt’s account suggests that they can; some suggests that they can’t. Second: what entities in the world actually qualify as wantons? On the one hand, Frankfurt suggests that the class of
wantons “includes all nonhuman animals that have desires and all very young children,” while excluding most adult human beings. Adult human beings may act wantonly, Frankfurt claims, but are not wantons tout court. This suggests that adult human beings are debarred from being wantons tout court. And yet Frankfurt’s paradigm examples of wantons are adult human beings who don’t merely act wantonly but are themselves wantons tout court. Further, his equivocation on the first question in this paragraph and his equivocations elsewhere as to whether wantons can reason make it difficult to exclude the possibility that an adult human being could be a wanton tout court. We can infer from this that both the status and the predicament of a Frankfuritian wanton is a fairly confused matter.

This confusion, I’m suggesting, finds its way into Korsgaard’s account, raising the troublesome issue: what’s the problem with being a wanton? Why is it so obviously a problem to be one? There are two possibilities here, neither of them helpful to Korsgaard.

The first is that it’s just impossible for a human adult to survive a normal lifespan as a wanton, in the way that it’s literally impossible to survive a lifespan without breathing. Presumably, the thought here is that some brute psycho-physiological mechanism just kicks in and forces one out of wantonhood in the way that one’s respiratory system forces one to breathe. But if this option is right, Korsgaard is misdescribing the phenomenology of reflection in a radical way. It is not that we have impulses that make us wantons, and then rise freely to reflection. It is that we have impulses that make us wantons—and other impulses that drive us out of wantonhood.

A second possibility is that while it’s possible to be a wanton, it is highly inadvisable to do so. A life lived as a wanton would likely or inevitably be highly
arduous, risky, or somehow subversive of physical or mental health—or perhaps just in some sense impoverished. So while it is in an agent’s power to remain a wanton, he shouldn’t. The trouble is, however plausible this view may be, we need an argument for it, and the minute we produce one, we see a major conflict between any plausible version of the argument and what Korsgaard has identified as the source of normativity.

In “Freedom of the Will and the Concept of a Person,” and elsewhere, Frankfurt seems to suggest that a wanton’s life would be a brutish, subhuman, animalistic one, and intimates that these features would make wantonhood something to avoid. But Frankfurt offers no explicit argument to that conclusion, and in any case, the most plausible argument to that conclusion conflicts with Korsgaard’s account of the source of normativity. An argument to the relevant conclusion would have to show that a subhuman (i.e., wanton) life is normatively inappropriate for adult human beings. This is tantamount to saying that adult human beings cannot flourish as wantons. But an argument that tells us that places the source of normativity in something like a conception of human flourishing, and offers no obvious place to reflection as the source of normativity. The point is that reflection makes some contribution to well-being in virtue of which wantonhood is subpar. It cannot be, after all, that wantonhood is subpar simply because wantons fail to reflect. That is simply a statement of the defining characteristic of wantonhood, not an account of what is wrong with being one.

There are a number of problems lurking here, but the most fundamental one is that Korsgaard has simply assumed that the fact that an agent faces alternate possibilities implies by itself that the agent is required to choose between them. This is just the
“capacity-to-requirement” argument of the previous section presented in a new guise. But the fact is,

(i) S’s facing the possibility of {φ-ing or ~φ-ing}

is not equivalent to

(ii) S’s facing the necessity of {φ-ing or ~φ-ing}.

Since the claim about possibility doesn’t entail the one about necessity, Korsgaard needs an argument that supports the move from (i) to (ii).

As I’ve been suggesting, the covert assumption seems to be a commitment to some quasi-Aristotelian conception of flourishing or a humanly appropriate life. In other words, the move from (i) to (ii) is mediated by something like the underlined clause in (iii):

(iii) S can either {φ or ~φ} and if S is to flourish, S must either {φ or ~φ},

for values of “φ” that make a difference to human flourishing (e.g., “being a wanton”). But that is neither what Korsgaard explicitly says nor what she is entitled by her theory to say.

The underlying difficulty here arises, in my view, from the somewhat uneasy relationship between Frankfurt’s and Korsgaard’s theories. Korsgaard tells us that her own account of normativity is “sketchy and sketchily argued.”36 It is, at the same time, an argument with strong avowed affinities with Frankfurt’s moral psychology and theory of volition. Given the sketchiness of the argument and the vagueness of the affinity, it is very easy to read Frankfurt’s argument wholesale into Korsgaard’s without noting the difference in context between them. But the context matters. Frankfurt’s “Concept of a
Person” essay is a piece of descriptive metaphysics in the Strawsonian mold, that aims to identify the volitional structure characteristic of persons. Korsgaard’s *Sources* is a piece of meta-ethics that seeks to identify the source of normativity. The one project is primarily descriptive, the other justificatory, and Korsgaard’s theory blurs the difference between those two aims.

To be more specific, Korsgaard, in relying on Frankfurt, conflates two things that Frankfurt himself often conflates:

1. The volitional structure *definitory* of a person, in Frankfurt’s sense.
2. The *normative requirement* that one strive to become a person in Frankfurt’s sense, and maintain that identity.

Claim (1) tells us what structure an agent’s volitions will have if, dissatisfied with being a wanton, that person actualizes her capacities to become a person. It does not say that one should want to be a person, nor does it give reasons for not wanting to be a wanton. Claim (2) is a normative account that takes claim (1) as settled, and then seeks to offer a normative justification for why one ought to actualize one’s capacities for personhood, rising above wantonhood. But while claim (2) presupposes (1), claim (1) cannot by itself underwrite claim (2): we need a distinct and explicit argument for it.

In fact, if we take Korsgaard absolutely literally—if we hold her strictly to her account of the conditions on the normative question (as discussed in Chapter 7)—she is tasked with explaining *to the satisfaction of a wanton* why the wanton should not want to remain one. But this aspiration is doubly problematic, first because she offers no such account, and second because it is incoherent. A wanton by definition has no resources for being discontented with her predicament, or of grasping alternatives to it. So a wanton
could not coherently be convinced of the inadvisability of being one. A wanton, qua wanton, has no concept of “inadvisability” at all.

If we then weaken Korsgaard’s address condition (as clearly we must), she has to explain why a wanton is obliged to become a person (or act as a person). But she can’t do this (as I think she tries) simply by assuming that personhood is the privileged perspective from which wantonhood is to be judged wanting (so to speak). That begs the question. An account of the source of normativity has the burden of explaining why Frankfurtian personhood is the perspective from which wantonhood is to be judged a state that an agent ought not to want to be in. And that Korsgaard has not given us.

8.6 Problem (4): The Method of Reflective Endorsement

I’ve so far argued that Korsgaard fails to account for the normativity of first-order desires, gives us no reason why the agent must reflect, and gives us no reason to think that if the agent does reflect, he faces what Korsgaard calls the normative problem. For purposes of this section, however, let’s set aside all three of those criticisms and imagine that Korsgaard’s argument was successful at each point. That gives us an agent with first-order desires who rises to reflection and faces a normative problem.

At this point, according to Korsgaard, the solution to the problem lies with reflection—i.e., with what in Lecture 2 she calls the method of reflective endorsement, which she describes as follows:

If the problem is that morality might not survive reflection, then the solution is that it might. If we find upon reflecting on the true moral theory that we still are inclined to endorse the claims that morality makes on us, then morality will be normative. I call this way of establishing normativity the ‘reflective endorsement' method.39
Though she finds the method at work in Hume, Mill, and Bernard Williams, the method's defining principle comes from Hume:

These facts suggest that Hume is relying on an account of normativity which is completely general, applying to any purportedly normative claim. Let me define two terms that will help express this view. Call a purportedly normative judgment a 'verdict' and the mental operation that gives rise to it a 'faculty'...According to this theory a faculty's verdicts are normative if the faculty meets the following test: \textit{when the faculty takes itself and its own operations for its object, it gives a positive verdict}.\textsuperscript{40}

But two obvious problems obtrude here. First, why think Korsgaard’s claim true? Second, why think it is determinate enough to solve the normative problem?

An initial and obvious difficulty is that Korsgaard's discussion of the method arises in the context of a largely interpretive discussion devoted principally to Hume, secondarily to Williams, and finally to Mill. The interpretive nature of the discussion precludes Korsgaard from giving a direct argument for the principle. And so at one level we might simply regard it as unmotivated.\textsuperscript{41}

If we set aside truth-conducivity as a rationale for the principle, we're pushed to what Korsgaard had earlier called her procedural realism. "The procedural moral realist," recall, "thinks that there are answers to moral questions because there are procedures for arriving at them."\textsuperscript{42} But correct in virtue of what? The answer seems to be, in virtue of the faculty's positive verdict on itself. But then, in virtue of what does the faculty offer such a verdict? We've seen that the answer can't be truth-conducivity, but it's unclear where the answer then lies. As far as I can see, Korsgaard seems to be saying that a faculty renders a positive verdict on itself when it follows procedures that lead to such a
verdict, and the procedures lead to the verdict because the faculty has given a positive verdict on itself. The principle and method, then, seem indeterminate, and the argument for it, viciously circular.

8.7 Notes

1 Thanks to Carrie-Ann Biondi, Teresa Celada, Hilary Persky, Jason Raibley, and especially Michael Young for helpful comments on this material.


9 The issues are discussed with what strikes me as remarkable unclarity in the literature, so that while the terminology I use corresponds to the standard terminology in the literature, my way of using the relevant terms may not.

The distinction in the text is often attributed to Thomas Nagel, who makes a distinction between "motivated" and "unmotivated" desires in Thomas Nagel, *The Possibility of Altruism*, (Princeton: Princeton University Press, 1970), pp. 29-30. Alluding somewhat vaguely to Aristotle's *Nicomachean Ethics* III.3, Nagel describes "motivated desires" as those "arrived at by decision and after deliberation," and "unmotivated desires" as those that "simply assail us" because they do not come to us after decision or deliberation.
I find Nagel's conception of an "unmotivated" desire highly unclear. There is first the problem of describing such desires as "assailants": it's unclear whether Nagel means that unmotivated desires assail qua unmotivated, or whether the assailing species of unmotivated desires are the most paradigmatically unmotivated. Unfortunately, the only example of an unmotivated desire that Nagel gives is "hunger…produced by lack of food." Hunger may indeed seem at times to "assail" us (though this seems a highly idiosyncratic account of the phenomenology, dependent, it would seem, on one's attitudes toward food), but hunger so conceived is not a desire but a sensation of deprivation; the relevant desire would be the desire to eat something. It is, however, unclear to me why the desire to eat should be understood as "unmotivated." Couldn't the desire have arisen as a consequence of very general decisions and deliberations about when it's right to eat—decisions and deliberations that produce the desire to eat at more or less regular times?

In his The View from Nowhere (Oxford: Oxford University Press, 1986), p. 151 n. 3, Thomas Nagel drops the reference to Aristotle, describing the motivated/unmotivated distinction as having "a good deal in common with Kant's distinction between inclination (Neigung) and interest (Interesse)," in Critique of Practical Reason. It's not clear whether Nagel sees the Aristotelian distinction and the Kantian one as the same. If they are different, then his view would appear to have changed between Possibility of Altruism and View from Nowhere—a fact he never acknowledges. But if they are intended as the same, it's unclear how Nagel takes himself to reconcile an Aristotelian moral psychology with a Kantian one. Here, Nagel describes motivated desires as those that appear "only because I recognize that there is reason to do or want something." Again, it is quite unclear (to me) how a desire can "appear" under any other circumstances, in which case the distinction between motivated and unmotivated is obscure.

R. Jay Wallace criticizes Nagel's view, in R. Jay Wallace, "How to Argue About Practical Reason," Mind, vol. 99 (July 1990), pp. 362-66, but in my view simply replicates the difficulties in it. "Nagel," Wallace writes, "attaches great significance to the distinction between motivated and unmotivated desires, without making it adequately clear what it is for a desire to be motivated" (p. 374). But that seems to me to get things backwards. It seems clear enough that desires can somehow be motivated; what needs explication is what it is for a desire to be unmotivated.


13 Korsgaard, *Sources of Normativity*, p. 91.

14 Korsgaard, *Sources of Normativity*, p. 149.

15 Actually, things are a little more complicated than this, since Korsgaard describes a living thing's ultimate end as self-maintenance and reproduction, and we might wonder how this can count as a single unified end in those cases where the requirements of the two compete. This is a complication for an Aristotelian account, but not in my view a fatal objection. For present purposes, we can ignore the complication and still accept the point made in the text, since the point I'm making is that the Aristotelian account she offers is at odds with the Kantian one however it's understood.


18 Korsgaard, *Sources of Normativity*, pp. 46 and 93.

19 She uses the term “unconditional.” See Korsgaard, *Sources of Normativity*, pp. 102-3.

20 Korsgaard, *Sources of Normativity*, p. 46.

21 Korsgaard, *Sources of Normativity*, p. 93.

22 Discussed above in Chapter 7.


24 Korsgaard, *Sources of Normativity*, pp. 91-97.

25 Korsgaard, *Sources of Normativity*, p. 98.


28 There is a separate problem here, noted both by Raymond Geuss (in Korsgaard, *Sources of Normativity*, p. 190) and by Simon Blackburn (*Ruling Passions*, pp. 252-56). Throughout *Sources of Normativity*, Korsgaard assumes that we are so passive with respect to the formation of our first-order desires that we’re obliged to say that we “find” ourselves with them and that they “dominate” us. Much of the plausibility of Korsgaard’s claims here turn on rhetoric of this sort, but if one doesn’t find the rhetoric plausible, or finds it overblown, the view loses much of its plausibility. Note that this rhetoric also trades on the equivocation between simple and complex first-order desires noted in section III. It is easier to accept Korsgaard’s rhetoric for simple than for complex first-order desires.

29 Frankfurt occasionally seems to be urging a principle to the effect that "If $S$ is a wanton at $t$, but generically capable of not being a wanton, $S$ has reason at $t$ to stop acting wantonly." But the claim is never made explicit, and to the extent that it is, Frankfurt either asserts it by fiat or offers egoistic considerations for $S$'s having reason not to act wantonly. Obviously, Frankfurt's sheer assertion doesn't help Korsgaard, but—given her rejection of egoism—neither do the egoistic considerations he adduces. See Frankfurt, "Necessity and Desire," p. 105, in Frankfurt, *Importance*, and Frankfurt "The Faintest Passion" and "On Caring," in harry G. Frankfurt, *Necessity, Volition, and Love* (Cambridge: Cambridge University Press, 1999), pp. 104 and 162, respectively. Frankfurt's view has the additional debility of at once affirming a form of egoism (in the passages just referenced) and contradicting it (see Frankfurt, "Autonomy, Necessity, and Love," pp. 133-35 and "On Caring," pp. 165 and 174, respectively, in *Necessity, Volition, and Love*). Thanks to Michael Young and Carrie-Ann Biondi for helpful suggestions on this issue.


relevant ambiguities in Frankfurt’s account. All page references are to Frankfurt, “Freedom of the Will.”

TABLE 8.1
AMBIGUITIES IN FRANKFURT’S ACCOUNT

<table>
<thead>
<tr>
<th>Case</th>
<th>What it is</th>
<th>Does it have first-order desires?</th>
<th>Does it have second-order desires?</th>
<th>Does it have second-order volitions?</th>
<th>Page reference in Frankfurt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>person (paradigm case)</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>p. 16</td>
</tr>
<tr>
<td>2</td>
<td>“higher” wanton</td>
<td>y</td>
<td>y</td>
<td>n</td>
<td>p. 16 n. 5</td>
</tr>
<tr>
<td>3</td>
<td>person (non-paradigm case)</td>
<td>y</td>
<td>n</td>
<td>y</td>
<td>pp. 15-16</td>
</tr>
<tr>
<td>4</td>
<td>“lower” wanton</td>
<td>y</td>
<td>n</td>
<td>n</td>
<td>p. 16 n. 5</td>
</tr>
<tr>
<td>5</td>
<td>person (non-paradigm case)</td>
<td>n</td>
<td>y</td>
<td>y</td>
<td>pp. 14-15</td>
</tr>
<tr>
<td>6</td>
<td>person (non-paradigm case)</td>
<td>n</td>
<td>y</td>
<td>n</td>
<td>pp. 14-15</td>
</tr>
<tr>
<td>7</td>
<td>person (non-paradigm case)</td>
<td>n</td>
<td>n</td>
<td>y</td>
<td>pp. 14-15</td>
</tr>
<tr>
<td>8</td>
<td>death, sleep, coma</td>
<td>n</td>
<td>n</td>
<td>n</td>
<td>Not explicitly discussed</td>
</tr>
</tbody>
</table>

34 See note 30 above.

35 Consider two of them. (1) Korsgaard assumes that reflection weakens the agent’s attachment to first-order desires. This assumption undergirds her view that in coming to reflect, the agent comes to face a quandary about what to do. But it’s not clear that reflection always weakens attachment; couldn’t it augment it? In other words, why assume that reflection always yields motivational distance (unless we’re simply reserving the word ‘reflection’ for that)? Couldn’t it just yield the reverse—a sort of motivational
proximity? If so, an agent might rise to reflection but never face a normative problem at all. Thanks to Hilary Persky for helpful discussion on this point.

(2) Second, Korsgaard seems to assume that reflection is necessarily a propositional phenomenon. This constitutes the basis of her confidence that in opening up motivational distance, reflection would necessarily express itself as a question. But this overlooks considerable ambiguity in the Frankfurtian (and ordinary) notions of self-consciousness and reflection. (a) There is a weak sense in which we—humans and higher mammals, certainly—can reflect non-propositionally: we can, for instance, direct our attention toward and away from objects of awareness. Directing attention away from such an object can produce what Korsgaard calls “distance” without doing so propositionally. (b) There is also a distinction to be drawn between reflection of a sort that leads the agent from first-order desires to second-order volitions, and reflection that operates within the first-order level. It takes a certain sort of reflection to get from “I want to eat” to “I want oysters as an appetizer,” but since the reflection takes place at the first-order level, it is clearly not what Korsgaard has in mind when she uses the term “reflection.” (c) There is a family of concepts for self-consciousness quite unrelated to what Korsgaard has in mind by either “self-consciousness” or “reflection.” Dolphins and small children recognize themselves in mirrors; this form of self-awareness is often described as “self-consciousness” without involving the full-blooded notion that Korsgaard describes in Sources of Normativity.

36 Korsgaard, Sources of Normativity, p. 91.


38 A wanton can be discontented in a situation, but on Frankfurt’s view, she can’t be discontented about her volitional predicament.

39 Korsgaard, Sources of Normativity, pp. 49-50.

40 Korsgaard, Sources of Normativity, p. 62 (Korsgaard's emphasis). This definition is not entirely clear as stated. Korsgaard defines a “faculty” as a type of “mental operation,” but then states a principle that describes faculties as something distinct from operations. A faculty, I assume, is a capacity for a mental operation. (The discussion of Hume, with references, is in Korsgaard, Sources of Normativity, pp. 50-66.)

41 An interpretive problem arises as well: Korsgaard ascribes the method to Hume, but it's not clear whether she means that (a) one can find elements of the method in Hume, or that (b) Hume was himself committed to the method as a method. Prima facie, reading (a) seems more plausible.
42 Korsgaard, *Sources of Normativity*, p. 36.
The framework of the preceding chapters puts us in a unique position to evaluate Korsgaard's project as well as kindred projects. In particular, it gives us a sense of what questions to ask, of the answers to them, and of a corresponding verdict on Korsgaard’s project.

First: If Korsgaard rejects foundationalism in epistemology, how does she hope—if at all—to reconcile her project with claims about the truth of moral propositions? As I argued in Chapter 7, the price of Korsgaard's rejection of epistemic foundationalism is a highly problematic rejection of moral realism, and with it, of moral truth in favor of a somewhat obscurely described and thinly defended form of “constructivism.”

Second: How does Korsgaard formulate what she takes to be the fundamental question motivating her meta-ethical project? Though more self-conscious than most, I've suggested in Chapter 7 that Korsgaard pays insufficient attention to the task of formulating what she calls "the normative question." As I argued previously in this study, the proper formulation of a question determines the overall structure of a theory by determining the further sequence of questions to be asked (or not asked), and by this expedient circumscribing the range of permissible and impermissible (or unthought-of) answers. By this measure (indeed, her own), Korsgaard fails to answer the fundamental question I formulated in Chapter 5 and answered in Chapter 6.
Third: *What* does Korsgaard take to lie at the foundations of ethics, and how does she establish it as the foundation? In Chapter 8, I argued that her argument fails, and that its failure flows more or less directly from the erotetic failures discussed in Chapter 7. Not having clearly formulated the foundational question, she fails to give an answer that clearly identifies a genuine foundation for ethics.

The weaknesses of Korsgaard’s theory do not, of course, constitute an argument for my view. But they do bring out some of its stronger features. Korsgaard’s theory fails, I suggest, insofar as it lacks precisely what I take my view to have: a proper formulation of the question motivating her meta-ethical foundationalism; a defensible account of the foundation itself; an account of the relation between the meta-ethics and some epistemology; and by implication, an account of the relation between the meta-ethics and moral truth. If I am right, a defensible version of Korsgaard’s account—of a Kantian ethic, generally—would have to develop those things. The question would then be which account, if fully developed, offered the more powerful brand of normative foundationalism. Though I don’t answer that question here, the issues I raise are a necessary preface to that discussion.
Rawls’s views on the relation between realism and constructivism, though highly relevant to Korsgaard’s project in *The Sources of Normativity*, are difficult to understand and also difficult to put together into a coherent picture. When Rawls speaks of moral realism, he does so by restricting his focus to “rational intuitionism,” treating it as the paradigm but not exclusive instance of the theory: “The first feature of rational intuitionism says that moral first principles and judgments, when correct, are true statements about an independent order of moral values; moreover, this order does not depend on, nor is it to be explained by, the activity of any actual (human) minds, including the activity of reason” (*Political Liberalism*, p. 91, hereafter *PL*).

Conceptions of moral realism might be differentiated by the meaning they assign to the notion of an “independent order of moral values,” but Rawls neither explicates the notion nor discusses non-intuitionist alternatives to realism. When he contrasts intuitionism with constructivism, it is therefore unclear how the contrast would work with respect to non-intuitionist versions of realism. But Rawls’s constructivism is in any case intended to be “limited to the political”; it is not meant to be a comprehensive moral doctrine, whether realist or constructivist (*Collected Papers*, 388n.2 [hereafter *CP*], 583,
Rawls thus goes out of his way to reject the idea that his political constructivism is to be equated with Kant’s moral constructivism, or Kantian moral constructivism generally.\(^1\)

Korsgaard’s references to Rawls are, in light of the preceding two points, rather puzzling. Korsgaard invokes Rawls’s constructivism as a paradigmatic form of procedural moral realism. But Rawls explicitly rejects realism and explicitly rejects the idea that his constructivism is a moral as opposed to political doctrine. While Rawls does claim—rather confusingly, in my view—that political constructivism is “compatible” with all “reasonable comprehensive doctrines,” realist, constructivist or otherwise (\textit{PL}, 90, 97-98, 100, 109, 113, 126-129), a theory’s mere compatibility with something \(X\) is hardly connection enough to make the theory a paradigm case of \(X\)—and less so, given the convoluted (or at least controversial) nature of the compatibility that Rawls alleges to obtain between “reasonable comprehensive doctrines” and political constructivism. It is thus difficult to see how Rawls’s view is a paradigm case of Korsgaard’s procedural realism.

But we can go further: even setting aside the preceding points, it is hard to see how Rawls’s view is in \textit{any} sense an instance of \textit{any} form of moral realism, procedural or otherwise. This becomes amply clear if we consider a survey, across time, of Rawls’s views on moral truth. There is first the fact that these claims seem to change across time, second their sheer unclarity, third (given unclarity) their apparent lack of diachronic coherence, and fourth their apparent rejection of the very claim Korsgaard ascribes to them.
In “The Independence of Moral Theory” (1975) Rawls tells us that his theory tells us to “bracket” the problem of moral truth, discussion of that topic being “premature” \((CP, 288, 290)\). In the same paper, however, he seems to suggest that the coherence achieved by the method of reflective equilibrium (if achievable) leads to moral truth, or at least a coherentist conception of truth according to which “the question as to the existence of objective moral truths seems to depend on the kind and extent of the agreement that would obtain among rational persons who have achieved, or sufficiently approached, wide reflective equilibrium. This illustrates the dependence of moral epistemology on moral theory” \((CP, 301)\).

In “Kantian Constructivism in Moral Theory” (1980), Rawls tells us three times that there are “no moral facts apart” from the procedure of construction envisioned by constructivism \((CP, 307, 350, 354)\). This is certainly consistent with Korsgaard’s ascription to Rawls of a rejection of substantive realism, but having denied the existence of moral facts, instead of denying the existence of moral truth, he once again brackets the issue, asserting that it is more “desirable” to speak of “reasonability” than truth \((CP, 340, 355; \text{cf.} 351-352)\). In the same place, however \((CP, 355)\), Rawls asserts that the preceding claim “does not imply that there are no natural uses for the notion of truth in moral reasoning,” a claim that Rawls neither illustrates nor explains and which at face value seems fairly mysterious in an author who has just gone out of his way to bracket the issue of moral truth and to deny the existence of moral facts.

Elsewhere, Rawls tells us that “a constructivist political conception is not at odds with our commonsense ideas of truth and matters of fact” \((PL, 122)\). But Rawls’s claims here are highly problematic. At times he seems to be affirming the existence of non-
moral or definitional truths about slavery. At times, he seems to be affirming the existence of moral truths, by way of a lapse into intuitionism (a common criticism of Rawls’s moral epistemology). At times what he says seems simply to contradict his rejection of moral realism. And at times it seems difficult if not impossible to decode the claims he makes.

In “Justice as Fairness” (1985), Rawls tells us, again, that claims to moral truth are to be avoided by the method of avoidance on the grounds that discussion of moral truth would be politically inexpedient (CP, 388, 394-5, 403n.22, 424, 434). But of course while justice as fairness does not depend—or hold itself out as depending—on claims to universal truth qua political conception, as we know from Political Liberalism, it is compatible with the (mutually incompatible) claims to universal truth made by other “reasonable” comprehensive doctrines.

In “The Idea of an Overlapping Consensus” (1987), Rawls insists that his political constructivism is neither indifferent to nor skeptical of moral truth (CP, 429, 434, with 434n.21), a claim echoed and discussed at some length in Political Liberalism (PL, 150-54). But it is frankly impossible to assess Rawls’s claims here given his previous insistence on “bracketing” and “avoiding” the very issue which he now claims is neither a matter of skepticism or indifference.

Finally, “The Idea of Public Reason Revised” (1997), re-asserts Rawls’s claim that moral truth is to be avoided in political discourse governed by “public reason” (defined as Rawls defines it) without denying its possibility, or expressing indifference or skepticism about it, while “replacing” it (CP, 574, 607, 608).
All things considered, it seems difficult to impose any interpretive order on the sheer number of claims that Rawls makes about moral truth. It is unclear what Rawls means by realism, what he means by moral truth, whether he takes truth to exist and in what sense, what he takes the truth-makers of moral truths to be, how he takes moral realism to relate to constructivism, or indeed, what constructivism is (in what sense claims can be constructed but not made and by what procedure, and what class of claims this applies to).

Notes

1 Though at one point, Rawls seems to backtrack in both directions, confining the scope of political constructivism to the use of “public reason,” i.e., “those questions in what I refer to as the public political forum,” so that political constructivism does not apply to “all political discussions of fundamental questions”—leaving open the possibility that such questions might be amenable to discussion in realist terms; see John Rawls, “Idea of Public Reason Revised,” in John Rawls, Collected Papers, ed. Samuel Freeman (Cambridge: Cambridge University Press, 1999), p. 575.


3 Rawls goes on to say in John Rawls, Political Liberalism (New York: Columbia University Press, 1993), that this compatibility entails that constructivist claims can, if supported by a true comprehensive doctrine, be made “correct” but not true (p. 128). It is unclear how to reconcile this latter claim with the assertion that moral truth is to be avoided and that justice as fairness does not depend on claims to it.
APPENDIX B:

RECONSTRUCTING KORSGAARD’S ARGUMENT: A CRITICAL SURVEY

Though Korsgaard’s argument in *The Sources of Normativity* has been widely discussed, it can reasonably be wondered whether the discussions correctly reconstruct her argument (and by implication, whether the structure of the argument is sufficiently clear to bear a rigorous reconstruction). Korsgaard describes her argument as “sketchy and sketchily argued,” but commentators have spent a remarkable amount of effort trying to reconstruct it, and their reconstructions have been surprisingly discrepant. One difficulty with almost all the commentaries, however, is how much of Korsgaard’s argument they grant without question. Given this agreement, or perceived agreement, commentators have more or less glossed over what I take to be the first few crucial moves of the argument, not only regarding them as uncontroversial, but not seeing that they are indeed distinct moves in the argument. In each case, I want to suggest that commentators have failed to grasp two essential points: (1) Korsgaard’s formulations in Lecture 3 fail to answer (or even to address) the most fundamental version of her own normative question, and (2) Korsgaard makes a subtle but unwarranted move from our *having a capacity for self-consciousness* to our *facing a normative problem*.
Allan Gibbard begins his discussion of Korsgaard by explicitly granting her starting points, and moving immediately to what follows from them:

Korsgaard’s starting points I find correct and illuminating; she presents ‘the normative problem’ as one that lies in wait for anyone who reflects and acts…My topic will be not whether her starting points are right or wrong, but what follows from them.¹

Thus Gibbard has nothing to say about Korsgaard’s formulation of the normative question in Lecture 1 or how that relates to the formulation of the normative problem in Lecture 3. He takes for granted that the two things coincide. But as I’ve argue in Chapter 7, they don’t, and their not doing so evident from his very way of describing Korsgaard’s argument. The most fundamental version of the normative question asks, “Why act?” Gibbard’s account of Korsgaard’s starting point assumes that the normative problem arises “for anyone who reflects and acts.” But then the question arises: why should anyone reflect or act? Since Gibbard’s version of Korsgaard’s normative question skirts that question, it cannot answer it. It therefore does not solve the normative problem that Korsgaard sets for herself.

John Searle frames his discussion of Korsgaard, somewhat oddly, in the context of a discussion of egoism and altruism—odd because Korsgaard is at pains to assert that there is in fact no genuine opposition between egoism and altruism. Egoism isn’t just false on her view, but unintelligible. In discussing this, Searle skips over the prefatory material in Lectures 1 and 2, and begins with Korsgaard’s attempt to derive the categorical imperative from freedom:

Let us then turn to examine Korsgaard’s Kantian account of how autonomy generates universality and universality generates altruism. Her solution is
presented as an interpretation of Kant’s views and here is how it goes: Kant argues that (1) we have to act under the presupposition of our own free will. He then continues that (2) free will, if it is to be a will at all, must be determined in accordance with a law. Since therefore, (3) free will has to be determined under its own law (by 1), it turns out that (4) the Categorical Imperative is a law of free will. The dubious step here is the second step. Why should the exercise of my free will in decision require any sort of law at all? Why can’t I freely decide what to do, just like that? Certainly no argument has so far been presented why there must be a law in order for me to make free rational decisions.²

In saying that “the dubious step is the second step,” Searle assumes that step (1) is obvious. But there are two controversial elements in step (1), one of which we can for present purposes set aside. The first is that when we act, we do so under the presupposition of our free will. The other, and for present purposes more important, is that we have to act. But Searle fails to see that no argument is given for this assumption—and none has been given despite the fact that Korsgaard herself raises the question of why we “have” to act. The fundamental problem, then, is not one that presupposes action, but one that explains why a free agent would, qua free, have to do anything. The problem that Searle identifies is surely ancillary to that one: Searle’s problem would not get off the ground until and unless the “why act” problem had first been solved.

J.B. Schneewind makes a similar mistake (page references in the block quote are Schneewind’s, and refer to Korsgaard’s Sources of Normativity):

Korsgaard’s non-reductive naturalized Kantian foundationalism (pp. 160-61) is a major contribution to current debates…Korsgaard wants to reply to the moral sceptic, understood as someone “who thinks that the explanation of moral concepts will be one that does not support the claims morality makes on us” (p. 13). The central question arises when an agent, deliberating about what to do now, asks whether morality’s claims are justified.³
One problem here is that the question that Schneewind puts in the mouth of the skeptic is not the same as he describes as “the central question”—a discrepancy he doesn’t explain. The second problem (and the deeper one) is that the skeptic is depicted as deliberating about what to do before we’ve answered the prior question of why the skeptic should act or deliberate on action at all. That fact suggests that the term “skepticism” is equivocal here: a weak skepticism might pose Schneewind’s question, but a stronger and more thoroughgoing skepticism would pose the “Why act?” question, thereby pushing Schneewind’s “central question” to the periphery.

In the Sources volume itself, G.A. Cohen offers a reconstruction of Korsgaard’s argument as follows:

1 Since we are reflective beings, we must act for reasons.

But 2 If we did not have a normative conception of our identities, we could have no reasons for actions.

So 3 We must have a normative conception of our identities (and our factual need for a normative identity is part of our normative identity).

So 4 We must endorse ourselves as valuable.

So 5 We must treat (all) human beings as valuable.

So 6 We find human beings to be valuable.

So 7 Human beings are valuable.

So 8 Moral obligation is established: it is founded in the nature of human agency.

The above argument can be decomposed into four subarguments, on each of which I now invite focus: (1) from 1 to 3; (2) from 3 to 5; (3) from 5 to 7; and (4) from 7 to 8.¹

Strictly speaking, the argument is not in standard form and thus is not, as stated, even a candidate for validity. Cohen contests the truth of (2) in an attempt to block the soundness of the first subargument, but given the face-value problems with validity, the
subargument wouldn’t have been sound even if the premise had been true. So for starters, I simply admit in all candor that I don’t see how Cohen’s reconstruction is supposed to work as a valid argument.

But suppose that it did work. The trouble is that Cohen leaves uncontested the most fundamental premise of the argument—premise (1). Why would the sheer fact of our being reflective beings—i.e., beings so constituted as to reflect if we choose to—entail that “must” act for reasons, or act at all? Why would a “can” of this sort entail anything of significance about any sort of “must”? And what sort of “must” is in play here. In leaving these questions unaddressed, Cohen ends up leaving Korsgaard’s basic argument essentially untouched.

Rachel Cohon reconstructs the argument as follows:

1. My reflective consciousness gives me a need for reasons to act and to live.
2. In order to have reasons, I need some practical conceptions of my identity.
3. Therefore, I have reason to have and be governed by some such conceptions of my identity.
4. This reason does not spring from the contingent practical identities I have. Rather, I have it in virtue of having reflective consciousness.
5. Therefore, in order to have any practical identities, I must value my reflective consciousness. (120-23).

Cohon skips a number between (4) and (6) to indicate what she takes to be a “gap” between those steps. But the interesting question, once again, is premise (1), and Cohon’s careful formulation of it brings the issue home. Premise (1) asserts that my reflective consciousness “gives” me the “need” for reasons to act and live. Notice that reflective consciousness is the source of normativity precisely in virtue of its giving me a need. But
how and why would my reflective consciousness “give” me any “needs” at all? In general, how and why do capacities give rise to needs?

We might try to make sense of (1) by assuming the general principle that

\[(CP, \text{ for ‘capacity principle’}) \text{ Any capacity (or any biological capacity, or psycho-biological capacity) generates its own needs.}\]

This is a bit of a stretch for any number of reasons,\(^6\) but grant it \textit{ex hypothesi}. The trouble is that if (CP) is true, its truth explains why (1) is the case. In other words, the truth of (CP) explains the connection between reflective consciousness and normativity. But in that case, wouldn’t (CP) be the source of normativity? And if some deeper principle explained (CP), wouldn’t that deeper principle be the source of normativity?

In pointing out the gap in Korsgaard’s argument, Cohon offers the following analogous argument, intended as a \textit{reductio}:

1’. My illness gives me a need for medicine.
2’. In order to have medicine, I need a doctor’s prescription.
3’. Therefore, I have reason to obtain a doctor’s prescription.
4’. This reason does not spring from any doctor’s prescription I have. Rather, I have it in virtue of having the illness.
6’. Therefore, in order to obtain a doctor’s prescription, I must value my illness.\(^7\)

Cohon’s point is that Korsgaard needs a step between (4) and (6) in the first argument to avoid the absurdity involved in (6’) in the second. I don’t happen to find Cohon’s reductio compelling: (6’) only seems absurd because (1’) is straightforwardly false. Claim (1’) only seems plausible (if it does) because it happens to be elliptical for the more awkward expression, “My need to cure my illness gives me a need for medicine.”
Here, “my need to cure my illness” is generated not by illness but by health. Unless we take (1’) elliptically in this way, (1’) is obviously false. Illnesses don’t generate needs; health does. It’s just that health generates very different needs when a person is ill (for medicine, prescriptions) than it does when the person is well (for neither). If (1’) is false, the argument is unsound and so its generating the absurdity involved in (6’) is irrelevant to Korsgaard’s argument.

What is interesting about Cohon’s argument is actually incidental to Cohon’s aim in presenting it. What the argument makes clear is that the most intuitive way of cashing out Korsgaard’s “must” is in terms of needs, but that very way of cashing it out reveals its fundamental defect. A capacity doesn’t generate needs except in the assumed context of a larger background of needs, and there is no such background unless we assume that the agent is already acting to satisfy his or her needs. Once we put the point that way, however, we find ourselves in normative territory that differs radically from what Korsgaard has presented us. The relevant issues are those discussed by an Aristotelian account of teleology, not a Kantian interpretation of laws. We need an account of the hierarchy of human needs, including the basic need that gives rise to needs as such.

I turn finally to Michael Bratman’s discussion of Korsgaard’s argument, one of the most complex of the various reconstructions of the argument in the literature. Despite its complexity, however, I don’t think Bratman circumvents the problem I’ve identified in the preceding reconstructions. In an initial overview of Korsgaard’s argument, Bratman writes:

Suppose I find myself faced with some desire or inclination to act in a certain way. As a reflective agent, I have the capacity to be reflectively aware of this
desire, \( D \), and of its support for action \( A \). Once I am reflectively aware I am faced with the question: “Should I act in the way \( D \) indicates?” To answer this query I need to decide whether or not to endorse acting on \( D \) in the present circumstance.\(^8\)

This description of Korsgaard’s argument bypasses why it is that the agent is required to ascend to the level of reflection. Bratman’s description is susceptible of a causal and a normative interpretation. On the causal interpretation, \( S \)’s possession of the capacity for reflection exerts causal pressure that more or less makes \( S \) reflect. On the normative interpretation, \( S \)’s capacity to reflect makes it possible for him to satisfy a requirement to do so. The causal interpretation is incompatible with the very point of Korsgaard’s argument, which is that we reflect freely. But if the normative interpretation is the right one, Bratman’s account doesn’t tell us where the relevant requirement comes from.

The phrase “As a reflective agent…” is a second source of ambiguity in Bratman’s interpretation. In one sense, it means “as an agent capable of reflecting but in no sense necessitated to do so.” Call this the capacity interpretation. In a second sense, it means “as an agent qua reflective.” Call this the exercise interpretation. If we adopt the capacity interpretation, we need an account of how to get from the capacity to the requirement. If we adopt the exercise interpretation, we begin with the existence of the requirement but have no account of how the capacity gives rise to it.

Bratman’s deeper analysis of Korsgaard’s argument suggests that he is offering a sort of hybrid of the causal and exercise versions of the argument. That is, he claims in effect that \( S \)’s possession of the capacity for reflection exerts causal pressure on \( S \) to reflect; this causal pressure causes \( S \) to reflect; hence \( S \) is a reflective agent insofar as \( S \) is caused to reflect.
The argument begins by supposing that

(1) I am throughout a fully reflective agent.

The idea behind this supposition is not that I am in fact always fully reflective, or given the press of events and the limits on our mental resources, should always be. We are interested in what is involved in being fully reflective because the capacity for full reflection is a basic feature of our agency. Fully reflective activity is the paradigm case of human action—it is human action “par excellence.” The argument will be that the pressure imposed by this paradigm of archetype is the source of practical normativity.

We now place the reflective agent in a circumstance calling for practical reflection. We suppose that

(2) I am faced with desire $D$ in favor of my now performing action $A$.

Since I am fully reflective I am reflectively aware that (2). This is not like being aware of a sudden urge to sneeze; for I see my potential action on the basis of $D$ as an exercise of my agency, and not merely the output of a causal process. Given that I see my acting on $D$ as a cause of my agency, I am faced with the question of whether to endorse my so acting. That is:

(3) I am faced with the questions: Should I now act on $D$? Should I endorse my now acting on $D$?

In fully reflective agency I act on $D$ only if I endorse my so acting.\(^9\)

I find Bratman’s account obscure. First, (1) defaults on the basic issue. It simply assumes that the agent is reflective; it doesn’t explain why she ought to reflect or has reason to.

What Bratman has done is to take the paradigmatic instance of human action and pick out full reflectiveness as one of its essential features. This is a description of that sort of action, not an account that explains why we have reason to perform it. Having bypassed the fundamental normative task, the move from (1) to (3) is easy: assuming that the agent has desires, and the agent is reflective in the requisite sense, the agent is faced with the “normative problem” of what to do. The trouble is that that “problem” has been front-loaded into the very concept of “a fully reflective agent”: a fully reflective agent just is an agent who, confronted with a first-order desire, regards that fact as a “problem.” “Full
reflectiveness” as a requirement has not been accounted for at all; it has merely arisen, causally, from the agent’s capacity to reflect. So what we are being told here is that if $S$ causally finds himself in a state which by stipulation is one in which he regards his own first-order desires as posing a problem, then $S$ faces a problem. That may be true, but it is not at all an account of the source of normativity.

Given the preceding, I’m inclined to agree with Wendy Donner: while “Korsgaard arguments are complex, interesting, and elegantly expressed,” she remarks, significant parts of it are “murky and not well explained or defended.”¹⁰ Indeed, it seems to me that the least well defended part of the argument is the part most in need of a defense.

Notes

¹ Allan Gibbard, “Morality as Consistency in Living: Korsgaard’s Kantian Lectures,” *Ethics*, vol. 110, no. 3 (October 1999), pp. 140-41.


Three reasons: (1) It’s not clear that CP is in fact true; (2) even if it’s true, it’s not clear it applies to reflective consciousness; and (3) it’s not clear that Korsgaard would endorse it.

Cohon, “Roots of Reasons,” pp. 75-76. To avoid confusion, I have added the apostrophe (’') in each case after the numbers of Cohon’s second argument.


Bratman, Review of Sources of Normativity, p. 701.

BIBLIOGRAPHY


Donner, Wendy, Review of *The Sources of Normativity*, *Dialogue*, vol. 36, no. 3 (Summer 1999), pp. 653-56.


Foot, Philippa, “Morality as a System of Hypothetical Imperatives” *Philosophical Review*, vol. 81, no. 3 (July 1972), pp. 305-16


Khawaja, Irfan, Review of Stephen Darwall’s *Consequentialism*, *Teaching Philosophy*, vol. 28, no. 3 (September 2005), pp. 281-84.


Smith, Tara, *Viable Values: A Study of Life as the Root and Reward of Morality* (Lanham, MD: Rowman and Littlefield, 2000).


