PRAGMATISM, GENEALOGY, AND MORAL STATUS

by

PAUL SHOWLER

A DISSERTATION

Presented to the Department of Philosophy
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2022

DISSERTATION APPROVAL PAGE

Student: Paul Showler

Title: Pragmatism, Genealogy, and Moral Status

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Philosophy by:

Colin Koopman	Chairperson
Mark Johnson	Core Member
Nicolae Morar	Core Member
Kristen Bell	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2022

DISSERTATION ABSTRACT

Paul Showler

Doctor of Philosophy

Department of Philosophy

June 2022

Title: Pragmatism, Genealogy, and Moral Status

This dissertation draws from recent work in pragmatism and philosophical genealogy to develop and defend a new approach for thinking about the concept of moral status. My project has two main aims. First, I argue that Huw Price's recent theory of philosophical naturalism, *subject naturalism*, can avoid several challenges by looking to the resources of philosophical genealogy, especially as it is developed in the work of Bernard Williams. Second, employing the methodological insights gained from this amended version of Price's project, I defend a genealogical account of moral status. Rather than theorize the grounds of moral status on the basis of an individual's properties or provide a conceptual analysis of *moral status*, my starting point is to look to the function that the concept plays within moral practice. In particular, I argue that it plays an indispensable, but overlooked role in allowing agents to deliberate about their practical identities and to articulate conceptions of moral progress. Taking this "function-first" approach, I argue, not only sheds light on various theoretical disagreements within applied ethics, but it advances debates concerning political and legal projects of affording rights to non-human animals, the natural environment (e.g., ecosystems), and machines displaying intelligence.

ACKNOWLEDGMENTS

I would like to thank my advisor, Colin Koopman, for his encouragement and incisive comments at every stage of the project. His intellectual curiosity and philosophical pluralism have inspired me more than I can say. Mark Johnson, Nicolae Morar, and Kristen Bell provided keen insights that undoubtedly strengthened my thinking. I am grateful to members of the Critical Genealogies Collaboratory, Brooke Burns, Colin Koopman, Mary McLevey, Zeinab Nobowati, Chris Shambaugh, and Valérie Simon, for their intellectual engagement with my research this past year. The Philosophy Department at the University of Oregon was, in many ways, the ideal place for me to carry out this work. I am especially indebted to Brooke Burns, Devin Fitzpatrick, Juan Sebastián Ospina, Claire Pickard, Rebekah Sinclair, and Kirstin Waldkoenig for their enriching conversations. In addition, I would like to thank the Social Sciences and Humanities Research Council of Canada for providing financial support that enabled me to complete this research.

For my parents, Lorie and Art.

TABLE OF CONTENTS

Chapter Page

CHAPTER 1

INTRODUCTION: PRAGMATISM AND METAPHILOSOPHICAL DISAGREEMENT

Longstanding, persistent disagreements are a pervasive feature of contemporary philosophy. Consider, for example, debates between *internalists* and *externalists* in epistemology, *reductionists* and *non-reductionists* in the philosophy of mind, or *realists* and *anti-realists* in metaphysics. In these kinds of disputes, new theoretical real estate is hard to come by. Tweaking or tinkering with established positions is always an option. But doing so is as exciting as it is likely to resolve the philosophical impasse. It is always possible to issue new proposals that challenge accepted assumptions; yet radical attempts to move beyond longstanding debates are most often greeted with suspicion and are viewed as attempts to change the crucial questions rather than answer them.

As a description of contemporary philosophy in its entirety, this is arguably a caricature. Nonetheless, it brings into focus two types of philosophical disagreement, which can be described in Kuhnian terms of competing paradigms. First, most persistent philosophical debates—such as the ones just mentioned—involve *intra-paradigm* disagreements that show no sign of abating. While, for instance, moral realists and anti-realists (at least appear to) agree about the terms of the debate, it is difficult to conceive of a consensus ever emerging from it. But most writers do not see this as a problem. Like Kuhnian practitioners of normal science they are usually content to continue their puzzle-solving. A second dispute, however, involves *inter-*

paradigm disagreements between those who operating within the confines of well-established views and the philosophical revolutionaries who hope to fundamentally alter the terms of the debate. The former think that we ought to continue searching for solutions to philosophical problems within the paradigm, whereas the latter see this task as misguided.

Both types of philosophical disagreement raise an important set of metaphilosophical questions. First, when it comes to the *intra-paradigm* disputes, one can ask: what, if anything, follows from the fact that such debates appear so intractable? What is the status of an intractable philosophical impasse? When, if ever, does it follow from the fact that a philosophical problem seems *irresolvable* that it ought to be set aside? The second set of metaphilosophical questions concerns *inter-paradigm* disagreement. In rejecting or resisting revolutionary proposals, those working within an established philosophical paradigm must appeal to a set of rational criteria on which to premise their rejection or resistance.[1] The problem is that those operating outside of the paradigm do not accept these criteria. In fact, they are trying to rethink or replace them. What is the status of these *second-order* philosophical disagreements? How ought they be approached?

One guiding idea of this dissertation is that while both types of metaphilosophical issues are a pervasive feature of contemporary philosophy, they are too often neglected. Indeed, as I shall suggest, this dissertation engages with two longstanding debates within contemporary philosophy which suggest them.[2] A second guiding idea is that *pragmatism* is the philosophical orientation best suited to deal with the kinds of theoretical impasses exemplified by both *intra-paradigm* and *inter-paradigm* disagreements. One characteristic of pragmatism is its high degree

[1] These criteria involve, for example, assumptions about the problems that a theory must resolve, constraints on an acceptable solution to those problems, among other things.

[2] The first is a debate in metaphysics concerning *philosophical naturalism*, and the second is a debate in value theory and applied ethics, concerning *moral status*.

of metaphilosophical self-consciousness.[3] On the other hand, pragmatism is a philosophical perspective that takes pluralism seriously. This is an asset when trying to navigate and shed light on disagreements between those who seldom (if ever) see eye-to-eye.

Like any philosophical tradition, however, pragmatism involves a complex, internally diverse set of ideas, methods, and commitments. So, I need to say more about what I mean by pragmatism. Before doing so, allow me to provide a brief statement of my aims and an overview of the following chapters.

My dissertation has two overarching aims. The goal of Part One (Chapters Two through Four) is to develop and defend a form of pragmatism that satisfies the demands of philosophical naturalism, while delivering a compelling account of normativity. I arrive at such a framework by putting into conversation Huw Price's account of philosophical naturalism, "subject naturalism", with the methodological resources of philosophical genealogy. Part Two (Chapters Five through Seven) employs the methodological insights gained from this combination of pragmatism and genealogy to develop and defend a novel theory of moral status, which I call a *pragmatic genealogy of moral status*. This approach promises to shed light on some of the central debates in recent value theory and applied ethics.

The aim of Chapter Two is twofold: first, I provide a detailed overview of Huw Price's *subject naturalism* (or "global expressivism") and locate it within a set of debates in contemporary metaphysics and the philosophy of language. Second, I outline the major philosophical objections to Huw Price's pragmatist view. At its core, subject naturalism inquires into philosophical topics (for example, *truth*, *meaning*, *value*, *mental phenomena*, or *modality*)

[3] This is true not only of recent pragmatists such as Richard Rorty, Hilary Putnam, and Huw Price, but metaphilosophical reflection is a hallmark of the writings of C. S. Peirce, William James, and John Dewey.

by turning our attention to the functions and geneses of our concepts and linguistic practices. Rather than ask about the essence or nature of moral facts or probabilities (to name a few examples), subject naturalists ask: how is it that humans came to employ those concepts within their social practices? What purposes do they serve for natural creatures like us? While I accept the general orientation of Price's pragmatism, I argue that, in its current form, it encounters two families of objections. First, critics have argued that subject naturalism is *normatively problematic*. This is because it relies on a detached, "third-personal" explanatory stance which threatens to undermine commitments to the very phenomena it seeks to explain. Second, others have argued that the subject naturalist's account of truth is *explanatorily inadequate*: it cannot allow for a distinction between epistemically virtuous practices and agonistic persuasive practices, it fails to explain why certain forms of inquiry (i.e., the sciences) involve a commitment to fallibilism, and it renders unintelligible the phenomenon of sociolinguistic change.

In Chapter Three, I advance an original argument, which shows how Price's account of truth fails to make sense of sociolinguistic revolutions (thus, further highlighting the explanatory inadequacy of his theory). My proposal considers how radical knowledge claims—especially those advanced in morality and science—often initially appear, from the standpoint of standard usage, to be patently false. In many cases, sociolinguistic revolutions involve these *abnormal* uses of language gradually coming to be accepted or adopted by a linguistic community. I argue that Price's account of truth is incompatible with this model of sociolinguistic revolution because of his overly restrictive, ahistorical conception of the norms governing assertion.

In Chapter Four, I argue that subject naturalism can overcome these normative and explanatory challenges by embracing the methodological insights of philosophical genealogy, especially as it is developed in the work of Bernard Williams. Combining these two perspectives yields a compelling form of pragmatist philosophical naturalism with broad explanatory applicability, and that is equipped to make sense of normative phenomena. For Williams, a philosophical genealogy involves two stages. First, it constructs a "state-of-nature" story—an explanatory model depicting human beings with various generic needs, capacities, and interests, the purpose of which is to explain how a certain conceptual practice came about. Second, a philosophical genealogy requires a *de-idealizing* or *historicizing* component, which aims to show how the target of the state-of-nature explanation has come to be elaborated and integrated into increasingly complex, functionally diverse, and historically contingent social practices. My central claim is that Price's subject naturalism amounts to something like a state-of-nature genealogy, but lacks a crucial commitment to de-idealization and historicization. Recognizing this lacuna helps diagnose the weaknesses in Price's pragmatism and points the way toward overcoming them. On the one hand, by taking on the de-idealizing and historicizing aspects of genealogical explanations, subject naturalists acquire a way to vindicate our commitments to certain conceptual practices, by showing how those items have come to fit within local constellation of values and concerns. This provides a response to the charge that subject naturalism threatens to undermine or subvert our commitments. On the other hand, by tracing the historical developments and transformations of the norms and values underlying our practices, subject naturalists will be better positioned to make sense of the distinction between epistemically virtuous and agonist-persuasive practices, the importance of regional fallibilism, and most importantly, sociolinguistic transformation.

Part Two (Chapters Five through Seven) draws from this amended version of Price's project to develop a *pragmatic genealogy of moral status*.[4] To say that an entity has moral status means that its interests ought to be taken into consideration when we deliberate about what to do, or that it can be an appropriate target of moral obligations (Warren 1997). Many philosophers aim to identify the properties or relations that ground moral status in hopes of determining whether or to what extent we have moral obligations to fetuses, infants, human beings with serious cognitive disabilities, non-human animals, ecosystems, intelligent machines, and so on. In this respect, a correct theory of moral status is often regarded as the holy grail of applied ethics, as it would provide a mechanism for settling disagreements about the proper limits of moral concern.

In Chapter Five I argue that there are two outstanding problems which existing theories of moral status have failed to adequately solve. The first, which I call the *problem of intractability,* stems from a longstanding theoretical divide within the literature on moral status. On the one hand, orthodox accounts subscribe to a view called *moral individualism*, the idea that an entity's moral status is grounded in its intrinsic properties. On the other hand, a growing number of writers—influenced by Ludwig Wittgenstein—have rejected moral individualism in favor of more contextualist, relational approaches that inquire into the conditions of possibility for moral status ascription. I refer to this position as *moral humanism*. The core theoretical divergence between these two camps has not been adequately understood, a failure which has left many debates in applied ethics at an impasse. The second outstanding problem, which I call the *problem of eliminativism* is that the concept of moral status is an idle cog within moral theory

[4] In general, this proposal can be understood as an expressivist account of moral status—in the sense of *global expressivism* that Price develops. Sometimes I refer to the project as an expressivist account of moral status, at other times I refer to it as a pragmatic genealogy of moral status. I explain these terminological choices in Chapter Six.

and practice. Eliminativists argue that the very idea of moral status is useless or confusing, and so ought to be set aside.

In Chapter Six, I develop a genealogical approach to moral status which can address the problems detailed in Chapter Five. Rather than ground moral status on the basis of an individual's properties or begin by reflecting on the meaning of our concepts, my starting point is to look to the function that the concept of moral status plays within moral deliberation. Following the methodological insights gathered from Part One, I begin by advancing a state-of-nature model that explains how human beings could have come to employ the concept of moral status in the first place. I contend that the notion of moral status performs both a *generalizing function* within moral practice (serving as a convenient shorthand for our moral obligations *as a whole*), as well as a *target-identifying function* that allows moral agents give voice to their practical disagreements, thereby facilitating social cooperation. In line with the pragmatic genealogist's methodological advice, I then offer an historical de-idealizing of this state-of-nature model, by suggesting ways in which conceptual practices involving moral status have been elaborated and diversified. On the one hand, I contend that the notion of moral status has come to play an important role in how we think about our practical identities. On the other hand, the concept has come to take on a progress-articulating function, whereby it has become tethered to a conception of moral progress as expanding the limits of moral concern. By tracking these developments, a genealogical approach can highlight the functional diversity that the notion of moral status has taken on. This provides explanatory leverage for making sense of the apparently intractable disagreement between moral individualists and moral humanists. On my view, these positions represent distinct deliberative strategies that answer to different practical problems that we face as moral agents. Moreover, I argue that by drawing attention to these overlooked functions, a genealogical approach provides a direct response to the problem of eliminativism.

Chapter Seven applies this genealogical approach to recent debates in applied ethics about the moral status of social robots. On my view, the question of "robot rights" (as it is sometimes called) centers on two fundamental disagreements: a *first-order* disagreement concerning the properties that are supposed to ground robot rights, and a *second-order* disagreement (between moral individualists and moral humanists) concerning how inquiry into the moral status of social robots ought to proceed, and what its aims should be. I develop and defend a *hybrid view*, which consists of two main theses: First, when it comes to *first-order* disagreements about the moral status of social robots, one ought to take a contextual, pluralistic, and open-ended view of the properties that can ground status ascriptions. Second, when it comes to *second-order* disagreements between individualist and humanist approaches to the moral status of social robots, one ought to delegate the deliberative strategies marked by these two positions to different domains of our moral practices. I argue that both moral individualism and moral humanism represent valuable deliberative strategies that answer to distinct practical problems that we are likely to face. While this hybrid account requires important modifications of and constraints on both positions, my account demonstrates that—when it comes to making sense of our obligations to social robots—both approaches are practically indispensable.

I have suggested that pragmatism is a compelling philosophical orientation for the central questions considered in this dissertation, due (in part) to its sensitivity to metaphilosophical questions along with its commitment to pluralism. But given the multitude of pragmatist options, it might be asked, why look to Price's subject naturalism as a starting point?

One reason for this is that Price's work is already firmly situated within debates about philosophical naturalism, and therefore, represents a promising avenue of moving contemporary philosophy in the direction of pragmatism. Like James, Dewey, and Rorty, Price is an adept critic of representationalism and its associated objectivist metaphysics and foundationalist epistemology. Unlike Rorty, however, he is—at least on the face of it—more sanguine about the possibility that pragmatism can offer *positive* philosophical insights, as opposed to making only *negative* points. Put another way, Price is more reluctant than Rorty to go in for a quietist view. Unlike James and Dewey, Price's employs a philosophical vocabulary and set of argumentative strategies that are more readily legible to philosophers working within contemporary analytic philosophy. This is not to say that the former pragmatists do not have much to offer to these contemporary debates. In fact, one of the aims of this dissertation is to call attention to the ways in which James's and Dewey's respective pragmatisms *are* relevant to contemporary debates. But taking either of them as my point of departure would require significant interpretive and contextualizing work that goes beyond the aims of this project.

Reading Price through a pragmatist lens reveals where his subject naturalism is best supplemented by other pragmatist insights. It also serves to demonstrate the value of Price's work for other pragmatists. In fact, one of my central claims is that the kinds of functional and etiological inquiries subject naturalism recommends are anticipated in the writings of other pragmatists, most notable James, Dewey, and to some extent, Rorty. Putting these views into conversation draws attention to an important *methodological* feature of pragmatism, which has yet to be adequately explored.

CHAPTER 2

SUBJECT NATURALISM AND ITS DISCONTENTS

2.1 Introduction: Overview and Aims

Towards the end of Part I of the Third Book of *A Treatise of Human Nature,* David Hume asks whether we should expect to find the sources of our judgments about good and evil "in nature" or whether we must look for them "in some other origin" (Hume 1978/1740, 473). "Our answer to this question," Hume writes, "depends upon the definition of the word, Nature, than which there is none more ambiguous and equivocal" (474). This is a rather remarkable thing for Hume to have written about a word which occurs not only in the title but in also in the very first sentence of his book. At the very least, one might wonder why he would have decided to wait nearly five-hundred pages before attempting to disambiguate it. Nonetheless, there is some truth to his claim, even if it tends to ring a little hyperbolic. From the standpoint of contemporary philosophy, the term 'nature' appears just as thorny as it did to Hume over two and a half centuries ago. Nowhere is this more evident than in debates about *philosophical naturalism*.

Many, if not the vast majority of contemporary philosophers can be considered naturalists in that they deny the legitimacy of positing *supernatural* phenomena within philosophical explanations of ourselves and what goes on around us. This is one respect in which Hume himself is arguably the naturalist *par excellence*.[1] His works situate morality—and human values

[1] For an insightful and nuanced discussion of how Hume's secularist approach can be seen as the culmination of a set of controversies within 17th and 18th -century British moral philosophy, see Michael Gill's *The British Moralists on Human Nature and he Birth of Secular Ethics* (2006).

more generally—firmly within a broader understanding of basic human capacities, needs and dispositions. At the same time, philosophical naturalism raises questions about the relationship between philosophy and the sciences. Many, but again not all philosophers want their theories and explanations to at least be consistent with the results of the best sciences of the day. If one's philosophical theory of the mind, for instance, were to fly in the face of empirical research produced by the cognitive and neurosciences, one would need to either revise one's theory or be prepared to relinquish one's naturalist credentials. Although non-supernaturalism and scientific respectability are important components of naturalism, for many philosophers they are not the entire story.

Most of the leading research programs in contemporary philosophical naturalism are committed to a set of stronger epistemological and ontological theses: that the only things that exist are entities posited by mature scientific explanations, and that all knowledge about the natural world must come from the natural sciences. Those who want to defend these views typically see themselves as tasked with solving "placement problems"—philosophical puzzles about how certain phenomena such as meaning, the mind, values, probabilities and other mathematical entities, morality, and aesthetic qualities are supposed to find a place within a world whose ontological contents are constrained by what successful science says there is.

In recent years, however, there have emerged important pragmatist challenges to this philosophical naturalism of contemporary metaphysicians. Thinkers like Richard Rorty and Hilary Putnam, and more recently Huw Price and David Macarthur have attempted to reclaim the label of philosophical naturalism while setting aside the objectivist metaphysics and its associated theory of mind and language to which it has become wedded. On this pragmatist reconstruction of naturalism, the fundamental guiding insight is that all aspects of human thought and behavior, from our ability to moralize to our ability to make counterfactual or probabilistic judgments must be understood in light of the fact that we humans are natural creatures whose capacities sustain our ongoing interactions with complex social and natural environments. From the perspective of contemporary philosophy this is a radical view—but it is by no means a novel one. This is the philosophical naturalism of earlier pragmatists, most notably John Dewey.

This pragmatist conception of philosophical naturalism has its advantages. As we shall see, it offers a compelling way of setting aside the placement problems over which metaphysical naturalists have spilled so much ink. Moreover, it does not consign us to a kind of quietism whereby philosophers are required to be resolutely therapeutic, refraining from venturing positive theses. At the same time this conception of naturalism tends to draw its fair share of opposition. There are at least two reasons for this. On the one hand, naturalism is thought to result in an uncritical acceptance of what is given to experience. Critics, going back to at least Kant, have held out for a conception of philosophy that is able to provide a transcendental justification for our knowledge claims, including our scientific ones. On the other hand, some philosophers contend that there are certain phenomena—perhaps consciousness, or moral truths—that naturalism simply cannot account for. One can call these the *problem of transcendence* and the *problem of completeness*.

In the first part of this dissertation, I shall defend a position which draws substantially from one of the leading proponents of both pragmatism and philosophical naturalism: Huw Price. Like Dewey, Price wants philosophers to rethink what it means to regard human beings as continuous with the rest of the natural world. And like Dewey, he believes that doing so will have profound implications for philosophy. Much like his early twentieth century forebearers, Price's philosophical naturalism has not only been the target of anti-pragmatist ire, but it has been the target of the kinds of criticisms just mentioned.

I have two primary aims in this chapter. The first is to provide a detailed overview of Price's *subject naturalism* as well as his constructive philosophical vision known as *global expressivism*. Following Capps (2018, 72), I take Price's account to be making both a set of negative (or critical) and a set of positive (or constructive) claims. In Section 2.2, I expound the former, and in Section 2.3, I expound the latter. My second aim is to identify and briefly discuss what I take to be the two most pressing objections to Price's position: [1] it misconceives the relationship between naturalism and science and thereby either fails to explain or ends up undermining important normative commitments; and [2] Price's naturalistic account of truth and assertion fails to account for important aspects of linguistic practice, most notably, it fails to account for sociolinguistic change, regional fallibilism, and the distinction between epistemically virtuous and agonistic-persuasive practices.[2] By drawing attention to these potential problems in this chapter, I hope to set the stage for Chapters Three and Four, in which I expand on and propose solutions to them.

2.2 Global Expressivism: The Negative Side

Global expressivism's negative component aims to radically reconstruct two mainstays of philosophical commonsense: *philosophical naturalism* and *representationalism*. Naturalism as we have already seen, is a highly contested position among philosophers.[3] That being said, Price

[2] These two criticisms can be understood as instantiations of the problem of transcendence and the problem of completeness.

[3] For helpful discussions of the concept of naturalism see Mario De Caro and David Macarthur's two edited volumes, *Naturalism in Question* (2004) and *Naturalism and Normativity* (2010). Especially their introduction to the latter.

takes its most general (and perhaps least controversial) formulation to be the idea that natural science "properly constrains philosophy" in the sense that "[t]he concerns of the two disciplines are not simply disjointed, and science takes the lead when the two overlap" (Price 2013, 3). The second mainstay is *representationalism*, which is the philosophical idea that the central function of language and thought is to represent aspects of some external reality (24). In particular, representationalists hold that substantive (i.e., non-deflationary) semantic relations or properties, paradigmatically *reference* and *truth*, are requirements of any explanation of the content of thought and language.

Price's project of reconstructing naturalism and representationalism depends on an important distinction between two approaches to philosophical naturalism, *object* and *subject* naturalism. Most philosophical naturalists are object naturalists, meaning that they subscribe to the idea that all that exists is the world as science sees it, or to the idea that only science affords us genuine knowledge (Price 2013, 4-5). Price wants philosophers to embrace an alternative view called subject naturalism, according to which "philosophy needs to begin with what science tells us about ourselves" (5).[4] As we shall see, he thinks that this shift will necessarily involve giving up representationalism as well. Although I trust that these distinctions will become clearer below, it is worth mentioning now that a core feature of (and perhaps motivation for) subject naturalism is that it claims a much more modest view of science's importance within the rest of culture than object naturalism does.[5] That is, subject naturalists—like so-called 'liberal

[4] That is, "Science tells us that we humans are natural creatures, and if the claims and ambitions of philosophy conflict with this view, then philosophy needs to give way" (Price 2013, 5).

[5] As Price explains, "Object naturalism gives science not centre stage but the whole stage, taking scientific knowledge to be the only knowledge there is (at least in some sense). Subject naturalism suggests that science might properly take a more modest view of its own importance. It imagines a scientific discovery that science is not all there is—that science is just one thing among many that we do with 'representational' discourse…The story then has

naturalists'—are more inclined to view science as one important activity (or better, set of activities) among others, and to deny that it ought to play some "foundational" role with respect to other areas of human life.[6] As Michael Bacon puts it, "subject naturalism provides a way of reconciling the demands placed on philosophy by scientific knowledge without opting for a reductive 'scientism' in which the only things that count are the objects of scientific inquiry" (Bacon 2012, 151).

To motivate the idea that subject naturalism should come to replace object naturalism Price presents a two-stage argument. First, he argues that subject naturalism is prior to object naturalism, in the sense that the latter depends on the former for its validity (Price 2013, 6). Second, he contends that from the perspective of subject naturalism itself, it turns out that there are good reasons to be suspicious of object naturalism (6). Price calls these the *priority thesis* and the *invalidity thesis*, respectively. To establish the priority thesis, he begins by examining the kinds of philosophical problems to which object naturalism tends to give rise. Many contemporary philosophers aim to address what are sometimes called *placement problems*. These are problems about how important things like *meanings*, *mental phenomena*, *mathematical entities*, *modalities*, *moral phenomena*, along with other things like *probabilities, causes,* and *conditionals* are supposed to "fit" or be "placed" within the world as described by science (5). It

the following satisfying moral. If we do science better in philosophy, we'll be less inclined to think that science is all there is to do" (Price 2013, 21).

[6] For a discussion of the concept of liberal naturalism see Macarthur (2014; 2015). I discuss Macarthur's views below. But briefly, one can envision different 'degrees' of naturalism, beginning at one end of the spectrum with narrow or strict *scientific* naturalisms which maintain that the only things that exist are the entities posited by the natural sciences. These narrow views could, in turn, be broadened to include the postulates of the human or social sciences as well. Finally, these broadened scientific naturalisms could be broadened even further to include entities which are not properly the objects of *any* kind of scientific study (namely irreducible normative notions). This extremely broad picture, on which the 'natural' is not identified in any way with the scientific but only by its contrast to the 'supernatural' is what Macarthur means by liberal naturalism.

is worth noting that the term ‘naturalism’ is often taken to be coextensive with the philosophical activity of addressing these kinds of problems.[7]

From the perspective of object naturalism, placement problems acquire their distinctive character because of an apparent disproportion between the number of true statements, on the one hand, and “truthmakers” licensed by the natural sciences, on the other. [8] Here, metaethical debates between moral realists and anti-realists provide a classic example. Many philosophers begin with the intuition that, at least in terms of their syntax, moral claims appear to be truth-apt. But if, as good object naturalists they accept that the only facts which can *make* such claims true are those which fall within the ontological purview of the natural sciences, they will begin to feel puzzled about their intuitions. The ingenious creatures that they are, philosophers have developed a set of “placement strategies” aimed at correcting for the imbalance between true statements and truthmakers (Price 2013, 26-9).[9] One might, for instance, endorse a form of *reductionism*, and argue that, despite appearances, there is really an isomorphism between statements of the domain in question and naturalistically certified facts.[10] Or, one might try to

[7] In his discussion of “the naturalist turn” in 20th century philosophy, Brian Leiter describes naturalism in exactly this way (Leiter 2004).

[8] According to Price, the structure of placement problems resembles that of the “matching games” that one might find in a children’s book, in which a set of stickers on one page are to be placed in the correct place on a corresponding image on the opposing page. Whereas the challenge for the child is to “fit” stickers onto the correct place within a complex scene; for the object naturalist, the challenge involves “fitting” the kinds of statements endemic to some domain of language (e.g., moral language) within the corresponding aspect of the world as described by the natural sciences. As Price explains, “[f]or each statement, it seems natural to ask what *makes* it true—what fact in the world has precisely the ‘shape’ required to do the job. Matching true statements to the world seems a lot like matching stickers to the picture; and many problems in philosophy seem much like the problems the child faces when some stickers are hard to place” (Price 2013, 23).

[9] See also (Price 1992, 37-41). Here, Price carves up the terrain slightly differently, arguing for a precursor to global expressivism called “discourse pluralism.”

[10] In metaethics, one form of realism that is both *naturalistic* and *reductionist* is exemplified by Peter Railton, who aims to show that moral facts are reducible to natural facts. Railton’s “generic stratagem” for naturalist realism is to “postulate a realm of facts in virtue of the contribution they would make to the *a posteriori* explanation of certain features of our experience” (Railton 1986, 172). He begins by arguing that we can posit such a set of facts in order

"grow the pie" of available "facts" or truthmakers in a realist manner, either by arguing that science is committed to the existence of a greater number of facts than philosophers have typically thought, or by arguing for the existence of non-natural facts.[11] Finally, another set of responses take seriously the idea that the abundance of true statements is merely apparent. *Eliminativists* assert that since some areas of discourse are devoid of referents or facts, they should be excised from our everyday idioms. Similarly, *error theorists* argue that since some area of discourse ultimately lacks legitimate referents or truthmakers, its claims are systematically *false*.[12] In a less radical vein, one might enjoin a kind of *fictionalist* stance towards such areas of discourse, maintaining that while there are, for instance, not *really* any "moral facts," there is some clear benefit to talking as though there were, such that no philosophical revision of our ordinary talk is in order.[13]

to make sense of an agent's (non-moral) objective interests (175-6). Then he shows how a similar argument can be made for objective moral facts. See Miller (2003, chapter 9) for a discussion of Railton's views and their reception. Reductionism is, of course, not limited to metaethical versions of the placement problem. For instance, mind-brain identity theories would be a paradigmatic example in the philosophy of mind. Price himself mentions Frank Jackson's work in metaphysics as a general kind of reductionist strategy (Price 2013, 27).

[11] Again, sticking to the case of metaethics, the former position is exemplified by the so-called Cornell Realists (i.e., Nicholas Sturgeon, Richard Boyd and Geoffrey Sayer-McCord) whose non-reductivist version of cognitivism construes moral properties as, for example, constituted by, supervening upon or be multiply realized by non-moral properties, but irreducible to them (see Miller 2003, chapter 8). Price points to David Chalmers' work in the philosophy of mind as another example (Price 2013, 27). By contrast, various forms of *non-naturalism* argue that there are sui-generis moral facts that not reducible to natural facts. Here G. E. Moore's account is perhaps the most famous example (Moore 1962/1903, 6-17). Crucially, for Price, although non-naturalism argues for the existence of non-natural facts, it operates in the same "ontological or epistemic keys" as their naturalist counterparts (Price 2013, 7).

[12] J. L. Mackie's position is a classic example of an error theory. On his view moral claims are cognitive (i.e., they *do* express beliefs purporting to be true or false), but since there are simply *no moral facts* to *make* such claims true, they are all systematically false.

[13] Price's own stance towards truth is in some respects close to this position (Price 2003, 180-1). He maintains that there is a clear value in employing truth-talk, but whereas the fictionalist takes a decisively anti-realist stance, Price opts for a kind of metaphysical quietism regarding the entire realism versus anti-realism debate.

2.2.1 The Priority Thesis

Price's priority thesis is the idea that subject naturalism is prior to object naturalism, in the sense that the latter depends on the former for its validity. His argument for this claim takes the form of a dilemma which considers two possible ways in which the object naturalist's placement problems could be thought to arise. On the one hand, one might think that placement problems originate as questions about human linguistic usage. We start, that is, by noting that human beings employ certain concepts and then wonder how, given "a commitment to object naturalism… what these speakers are thereby talking or thinking *about* could be the kind of thing studied by science" (Price 2013, 7-8). Call this the *linguistic conception*. On the other hand—and this is what Price labels the *material conception*—one might think that placement problems simply begin as puzzles about the nature of the objects themselves; for instance, as problems about *values*, *meanings*, *causes,* and so on (8). Price's strategy is to claim that [1] the linguistic conception of the origin of placement problems requires assumptions that belong to the domain of subject naturalism. And, [2] that that material conception is not a viable possibility for naturalists of any stripe.

Consider [1] first. According to the linguistic conception, placement problems originate *ex hypothesi* as questions about, for instance, mathematical or modal *language*, but very quickly become questions about non-linguistic *objects*. That is, the object naturalist begins by noting that we use mathematical and modal language to say apparently true things; but then (somehow) begins to wonder about the sorts of *entities* or *facts* to which such language could possibly refer.[14] How is this shift possible? According to Price, the answer is that it ultimately depends on

[14] For example, they wonder about the nature of mathematical or modal objects or facts.

a commitment to representationalism. It is only by making use of substantial semantic relations such as reference or truth—what Price calls a “semantic ladder”—that the object naturalist can subtly transform a question about *language* or *terms* into an issue about the place of certain *objects* or *entities* in the world (Price 2013, 9). But if this diagnosis is correct, then Price has a compelling case for the priority thesis. After all, representationalism looks very much like a theoretical claim about language: an issue which would seem to fall within the purview of subject naturalism. As Price explains,

> Given a linguistic view of the placement issue…substantial, non-deflationary semantic notions turn out to play a critical theoretical role in the foundation of object naturalism. Without such notions, there can be no subsequent issue about the natural ‘place’ of entities such as meanings, causes, values and the like. Object naturalism thus rests on substantial assumptions about what we humans do with language—roughly, the assumption that substantial ‘word-world’ semantic relations are a part of the best scientific account of our use of the relevant terms (10).

This path to the priority thesis, of course, depends on the assumption that the linguistic conception is on the right track. Price offers several reasons intended to show that the other side of the dilemma, the material conception, is not really a live option within contemporary philosophy.

On the one hand, the fact that non-cognitivism remains a viable response to many placement problems suggests, to some extent, that the linguistic starting point to placement problems is unavoidable (Price 2013, 17). Non-cognitivists typically argue that, despite *appearing* descriptive, truth-apt or belief-expressing, some areas of discourse are better understood as performing some other function.[15] To simply insist on a material starting point

[15] For instance, a non-cognitivists about probabilistic claims might argue that instead of thinking of utterances such as “it will probably rain tomorrow” as expressing straightforward beliefs which, in turn are supposed to represent probabilistic facts, we should think of such claims as expressing a speaker’s subjective credences or dispositions to bet.

would be to completely ignore the non-cognitivist's major insight, which is that one could learn something from subject naturalistic "reflections on the things that we humans do with language" (17). On the other hand, as Price points out, arguably the dominant *modus operendi* of contemporary analytic metaphysics is to "do metaphysics in a semantic key." That is, it is common for metaphysicians to characterize their work in semantic terms, for instance, in terms of "truthmakers" or "semantic role realizers." But, Price thinks, to adopt this stance is effectively to assume the linguistic conception.[16]

Price's argument for the priority thesis then, can be put as follows. The only viable route to object naturalism (i.e., the linguistic conception of the origin of placement problems) involves a commitment to representationalism. And since this commitment is effectively a theoretical claim about our use of language, it ought to be evaluated from the perspective of what our best scientific theories tell us about ourselves (i.e., from the perspective of subject naturalism).

2.2.2 The Invalidity Thesis

Price's second major thesis is that from the perspective of subject naturalism, there are good reasons to be suspicious of representationalism, and hence, object naturalism. He offers two main arguments in support of this invalidity thesis: an argument based on the attractiveness of

[16] This stance is nicely captured in a passage near the end of Timothy Williamsons' "Past the Linguistic Turn?" where he writes: "Some contemporary metaphysicians appear to believe that they can safely ignore formal semantics and the philosophy of language because their interest is in a largely extra-mental reality. They resemble an astronomer who thinks that he can safely ignore the physics of telescopes because his interest is in the extra-terrestrial universe. In delicate matters, his attitude makes him all the more likely to project features of his telescope confusedly onto the stars beyond" (Williamson 2004, 128). One influential approach to metaphysics that *might* be construed as circumventing the linguistic starting point is the so-called "Canberra Plan", associated with Frank Jackson and David Lewis (Price claims to have coined this moniker). Price thinks that this approach faces a dilemma—either retain commitment to naturalism by ultimately reverting to the linguistic starting point, or avoid the linguistic starting point by cutting themselves off from naturalism entirely (Price 2013, 18-20).

semantic deflationism and an argument intended to call into the question the very coherence of object naturalism.

The first argument appeals to the availability of semantic deflationism as an alternative to representationalism. Semantic deflationism (sometimes called "minimalism") denotes a family of approaches to semantic terms such as truth and reference, according to which these conceptions have no essential nature and can play no explanatory role in our theorizing (Tebbon 2015, 6). [17] Unlike, for instance, accounts which attempt to elucidate the nature of truth in terms of "correspondence with the facts," "coherence among a set of beliefs or propositions" or "what would be indefeasible at the end of inquiry," most contemporary deflationists maintain that the extent of philosophical insight which we can hope to glean about truth is reflected in (what is often called) the equivalence scheme:

ES: 'p' is true if and only if p.

ES is meant to capture the intuition that there is no apparent difference between asserting that a statement is true and asserting that statement. This "transparency property" of truth sets it apart from most other predicates, which tend to alter the contents of statements when addended in similar ways.[18]

[17] Helpful discussions of deflationism can be found in (Blackburn 1998; Brandom 2009, chapter 6; Tebben 2015; Lynch 2015; Horwich 1998; Dreier 2004; and Williams 2002).

[18] For instance, although asserting that '*p'* is clearly different from asserting that '*p* is funny', as we have just observed, asserting that '*p*' is not clearly different from asserting that '*p* is true.' Simon Blackburn employs the somewhat amusing metaphor of "Ramsey's ladder" (after Frank Ramsey, coincidentally, there is also a type of ladder called a 'Ramsay Ladder') to demonstrate that other semantic predicates share this transparency property. Ramsey's ladder is comprised of different 'rungs' each (supposedly) representing a different semantic 'level.' If the bottom rung corresponds to the assertion '*p*', and the next that 'it is true that *p*', then the third might correspond to 'it is really true that *p'*. Perhaps an even further rung might correspond to 'it is totally and absolutely really true that *p*'. Often, Blackburn remarks, philosophers tend to climb Ramsey's ladder and "announce a better theoretical view from the top" without realizing that the ladder is actually lying horizontally on the ground (Blackburn 1998, 294-7).

But if the distinctive feature about ‘truth’ is simply that it allows speakers to *reassert* things, one might wonder how such a ‘redundant’ property could have found its way into natural language.[19] Many deflationists have responded to this concern by highlighting the *role* that the truth predicate plays within our discursive practices. As it turns out, although truth has no essential nature it is still a handy linguistic tool. For instance, the predicate ‘is true’ is often characterized as a device of disquotation or as a logical device of generalization. It allows a speaker to endorse what others have said, even when they do not know what the contents of such sayings might be. After all, it is sometimes perfectly legitimate for me to claim: ‘Everything that my friends say about me is true’ without actually knowing (or even being able to enumerate) the details of what they have said. Employing the truth predicate in such a way would allow me to commit myself to their claims.[20]

At this point, deflationism begins to look like a decisively subject naturalist position. Rather than inquire into the nature of *truth* (inviting questions about how such a property could “fit” within the world as described by science), the deflationist’s glance is turned toward the discursive practices of human beings. As Price explains, deflationism “offers a broadly scientific

[19] Early deflationary accounts of truth were sometimes called “redundancy theories.”

[20] Paul Horwich has been a longstanding proponent of deflationism. He offers a nice summary of this view: “the truth predicate exists solely for the sake of a certain logical need. On occasion we wish to adopt some attitude towards a proposition—for example, believing it, assuming it for the sake of argument, or desiring that it be the case—but find ourselves thwarted by ignorance of what exactly the proposition is. We might know it only as ‘what Oscar thinks’ or ‘Einstein's principle’; perhaps it was expressed, but not clearly or loudly enough, or in a language we don't understand; or—and this is especially common in logical and philosophical contexts—we may wish to cover infinitely many propositions (in the course of generalizing) and simply can't have all of them in mind. In such situations the concept of truth is invaluable. For it enables the construction of another proposition, intimately related to the one we can't identify, which is perfectly appropriate as the alternative object of our attitude” (Horwich 1998, 2-3).

hypothesis about what linguistic creatures like us ‘do’ with terms such as ‘true’ and ‘refers’—what role these terms play in our linguistic lives” (Price 2013, 11).[21]

Why does semantic deflationism threaten to undermine object naturalism? The most obvious reason is that it is a resolutely non-representationalist position. Maintaining that the nature of truth is *entirely* captured by ES, is to reject the idea that truth is a substantive relationship that holds between *words* and the *world*. But without such relations, as we have seen, the would-be object naturalist is left without access to a “semantic ladder” which plays a crucial role in their formulation of placement problems (Price 2013, 22). Deflationists need not settle “issues about the items at the “world’s end” of such relations” such as referents, facts, truthmakers or semantic realizers (Price 2009b, 263), hence its availability as a theoretical alternative threatens object naturalism.[22]

A related argument that Price makes challenges object naturalism’s very coherence. As we have seen, the object naturalist is committed to the view that non-deflationary semantic notions form part of our best scientific theory about language. At the same time, as naturalists,

[21] Despite the fact that Price endorses deflationism about semantic notions (especially as they offer a promising alternative to representational paradigms) his relationship to these approaches is complicated. Insofar as deflationists try to limit philosophical inquiry into the function that truth plays within discursive practice, Price applauds them. In fact, such approaches clearly embody the kind of subject naturalism he endorses (9). However, as I explain below in more detail, Price thinks that most forms of deflationism fail to capture the normative dimension that truth plays in our lives (See [Price 1988]).

[22] At this point, one might protest that object naturalism can still function even with deflated semantic notions. After all, as is often remarked, while the left-hand side of the ES is about language, the right-hand side is about the world. Doesn’t this suggest that it is perfectly legitimate to investigate the “world” side of semantic relations? Price contends that this response is seriously misguided (Price 2013, 9). From the perspective of semantic deflationism, the object naturalist’s penchant for talking about referents or truthmakers ends up committing a fallacy of equivocation between the use and the mention of an expression. To see why, notice that the truth predicate’s disquotational function allows a speaker to use an expression that has been mentioned. For instance, following Quine, to say that “’snow is white’ is true” is to make use of a mentioned sentence—thereby making the claim that snow is white. But, recalling Price’s arguments in favor of the linguistic starting point of placement problems, “if our original question was really about language, and we rephrase the issue in…semantic terms [i.e., in terms of referents or truthmakers] we’ve simply changed the subject. We haven’t traversed the semantic ladder but simply taken up a different issue” (Price 2013, 9).

they would seem to be committed to the empirical contingency of their own theoretical claims. That entails, as Price explains, that "for any given term or sentence, it must be to some extent an empirical matter whether, and if so to what that term refers; whether, and if so where, it has a truthmaker" (Price 2013, 13). And while it might be possible for the object naturalist to adopt this stance towards *some* topics of investigation (for instance, they might be able to claim that whether mathematical terms have *referents* is somehow a contingent matter), "it seems impossible to make sense of this attitude with respect to the semantic terms themselves" (13). Price's point is that any object naturalistic investigation into whether 'truth' or 'reference' themselves have truthmakers or referents would necessarily *presuppose* the existence of the very semantic terms they would be purporting to investigate. That is, it seems impossible to ask whether the semantic relation 'reference' *refers* without making use of the semantic relation itself.[23]

So far, we have seen the general thrust of Price's criticisms of two fundamental assumptions in contemporary philosophy—object naturalism and representationalism. Taken together, they generate what seem like inescapable metaphysical problems about how certain phenomena are supposed to fit or be placed within the world as understood by science. Price is explicit that these considerations are not intended to be refutations, but rather attempts to show that much philosophical practice is not faithful to its naturalistic self-image. That being said, it is hard to overstate just how radical the critical component of Price's view is. If his arguments against object naturalism and representationalism are cogent, this would call into question some

[23] One option for the object naturalist at this point would be to simply insist that the existence of substantive semantic relations and properties is knowable *a priori*. Price could then object that this move undercuts any claim to naturalism. Moreover, it is worth mentioning that, as we have seen, semantic *deflationism* not only avoids this problem in the trivial sense that it makes no explanatory use of truth or reference, but it provides a completely respectable *naturalistic* stance towards those terms: effectively providing functional or genetic explanations of their use.

of the most fundamental assumptions guiding a significant number of directions of contemporary philosophy.

At this point someone might raise the following objection: doesn't giving up on object naturalism and representationalism mean the end of philosophical investigation into a wide range of important topics? Is Price just suggesting that we should turn our backs on the project of making sense of how, for instance, meanings, minds, morality and mathematical entities fit within our broader understanding of the natural world? Such a philosophical quietism is not a position that Price is willing to endorse. And to see how he manages to do avoid it; we will need to turn to the positive dimension of his global expressivism. This project involves a general departure from the kinds of questions motivated by object naturalism, towards forms of philosophical inquiry that are more self-consciously subject naturalist in spirit: a shift away from semantically-driven, metaphysical approaches to placement problems, towards what Price sometimes refers to as "linguistic anthropology" or "philosophical genealogy."[24] As he puts it, from this new perspective on the topics which have dogged analytic philosophers for the past century:

> The challenge is now simply to explain in naturalistic terms how creatures like us come to talk in these various ways. This is a matter of explaining what role the different language games play in our lives—what differences there are between the functions of

[24] As we shall see, Price does not distinguish between these two labels, nor does he offer a detailed explanation of what they ought to involve in terms of methodology. In discussing the "biological functions of the mental states we call beliefs", Price offers a gloss on the kinds of questions involved in linguistic anthropology or genealogy. He asks:

> How did it serve our ancestors to develop a psychology rich enough to contain such mental states? What role did those states play in increasingly complex lives? It is plausible, in my view, that there is no single answer, appropriate for all kinds of beliefs. Perhaps the function of some beliefs can be understood in terms of the idea that it is useful to have metal states designed to co-vary with certain environmental conditions, but for many, the story is surely more complicated. Consider causal or probabilistic beliefs, for example, which manifest themselves as dispositions to have certain sorts of expectations (and hence to make certain sorts of decisions) in certain sorts of circumstances. Plausibly, there's a [sic] interesting story to be told about the biological value of having internal functional organization rich enough to contain such dispositions (Price 2009a, 271).

> talk of value and the functions of talk of electrons, for example. This certainly requires a plurality in the world, but of a familiar kind, in a familiar place. Nobody expects human behavior to be anything other than highly complex. Without representationalism, the joints between topics remain joints between kinds of behavior, and don't need to be mirrored in ontology of any other kind (Price 2013, 20).

But before looking at the details of this proposal, I want to examine a set of concerns that several commentators have raised in relation to the scope of Price's naturalism. As we shall see, I believe that these objections will necessitate some revisions of Price's project. At the same time, given that they are complex and far reaching issues my aim in the next section is simply to explain them. The task of showing how Price's subject naturalism can overcome these worries is one I undertake in the fourth chapter.

2.2.3 Two Problems for Naturalized Normativity: Explanation and Commitment

Subject naturalism can be thought of as a kind of successor subject to the semantically-driven metaphysical projects comprising the more orthodox object naturalism. To succeed, it must be able to deliver satisfying naturalistic explanations of the various phenomena—for instance, minds, meaning, morality and mathematical entities—that have troubled object naturalists. By Price's own lights, many of these phenomena are *normative* phenomena, meaning that they are subject to standards of correctness and evaluable in normative terms. Consequently, it would be a serious problem if there were some features of the subject naturalistic perspective itself that precluded it from adequately accounting for these dimensions of linguistic practice. It turns out that several commentators have raised suspicions of its ability to deliver on this score. In this section I shall summarize two closely related problems of normativity that Price encounters.

The first worry is that subject naturalism cannot explain certain normative phenomena such as meanings, values and reasons. This difficulty, which stems from the fact that Price places too great of an emphasis on the connection between *naturalism* and *science*, pressures him towards a broadened conception of "the natural." The second worry has more to do with the normative implications of subject naturalism. Here the idea is not so much that it fails to make sense of normative phenomena, as that the very theoretical perspective of naturalism (in Price's case, this would mean genealogical or anthropological explanations of aspects of linguistic practice) may undermine or delegitimize normative commitments themselves. Of these two worries I take the second to be more threatening to Price's project. Although, as we shall see, it is also more difficult to articulate.

Like Price, David Macarthur has written extensively on naturalism in contemporary metaphysics.[25] He shares both Price's suspicion of the inevitability of philosophical placement problems, as well as his diagnosis of their origins in a misconceived conception of the scientific image. From the standpoint of an empirically informed conception of scientific practice, there is simply no reason to restrict ontological commitments in the way that orthodox naturalists typically have, and so there is no need to engage in kinds of metaphysical pursuits with which they have so frequently been occupied (Macarthur 2010; 2014). And although Macarthur has applauded Price's subject naturalism (along with Robert Brandom's normative pragmatics) as a promising way of arriving at a pragmatist account of normativity, he has more recently come to criticize Price for conceding too much to the orthodox conception: namely in his retention of the idea that naturalism is inextricably bound to the scientific image (Macarthur 2014). Whereas many commentators take Price's naturalism to be liberal in stripe—urging philosophers away

[25] See De Caro and Macarthur (2010), Macarthur (2008; 2010; and 2014) and Macarthur and Price (2009).

from the idea that "science should be given center stage"—Macarthur points to passages which belie the former's apparent ambitions to rid himself of the vestiges of scientism. For Macarthur, the only defensible form of naturalism is one that contrasts the natural, not with the non-scientific, but with the supernatural. That is, naturalists—as Macarthur sees it—should stick to eschewing notions like God's will, Platonic forms, Cartesian immaterial substance, and so on. By retaining a close methodological connection between naturalism and science, Price finds himself in an awkward position when it comes to his claim to offer a naturalistic treatment of normativity.

To see why, recall Price's dictum, that "philosophy should begin with what science tells us about ourselves" (Price 2013, 5). This may be reasonable advice when it comes to some questions. For instance, philosophers who want to know more about the nature of the mind would do well to pay attention to the results of recent neuroscience. But this advice seems far from helpful in other cases. The problem, as Macarthur puts it, is that "[s]cience has not shown that persons qua rational agents are fully understandable, or completely explicable, in scientific terms" (Macarthur 2014, 73). Human reason, history, art, and meaning are all phenomena that have yet to be the objects of successful scientific theorization (74).[26] What appears particularly troubling for Price is that this is also true of many aspects of language. The subject naturalist's *modus operendi* is to inquire—in the manner of a biologist or linguistic anthropologist—into the function or genesis of certain speech acts or concepts (Macarthur 2014, 76). Yet such an inquiry, Macarthur points out, is ambiguous. Is the subject naturalist supposed to study *meaningful utterances* (i.e., semantic, intentional, and normative linguistic expressions) or is she supposed to

[26] As Macarthur explains, "All of these only become fitting objects of scientific study if we re-conceive them according to concepts whose conditions of individuation accord with appropriate scientific standards of impersonality and determinacy. For example, a person qua living body can, of course, be studied by a physiologist or biochemist" (Macarthur 2014, 74).

study language construed as marks and noises? That Price includes locating “meaning facts” as a paradigmatic placement problem suggests that he has the former in mind (Macarthur 2014, 77). But is language (*qua* meaningful utterances) the sort of thing that can be adequately understood through scientific inquiry alone? Macarthur argues that it is not. Any meaningful utterance, he explains, “is a natural non-scientific item, which is not amenable (so construed) to naturalistic study” (Macarthur 2014, 77).[27] Following Quine, he contends that “meanings are too vague, interest-relative and subjective to be fit objects of scientific study. But it is meaningful language in this sense that is the supposed object of subject naturalist inquiry” (77). Indeed, the belief that meaning somehow *must* be open to scientific inquiry has given rise to one of the most intractable placement problems—what Macarthur calls “the problem of conceptual normativity” (Macarthur 2015, 571). As he explains, the problem is that “intentional content is constitutively caught up in various forms of normative assessment” which are not themselves amenable to scientific explanation and understanding (Macarthur 2015, 572). This is because the normativity of content has an irreducible first-personal dimension (beliefs necessarily involve normative statuses like *commitment* or *responsibility*), whereas the methods of the natural sciences are “distinctively third-personal” (575). The latter involve standards of objectivity and reproducibility that are in some sense incompatible with the first-personal standpoint. Any naturalistic inquiry of the sort

[27] There are at least two ways of interpreting Macarthur’s remarks here. On a strong reading, they can be taken to mean that *all* language *qua* meaningful expression is not the sort of thing that can be understood scientifically. A weaker reading would be that linguistic meaning is not the kind of thing that can be *exhaustively* understood through a scientific perspective (i.e., that anyone who claimed to be able to give a comprehensive account of linguistic meaning *solely* on the basis of a natural scientific explanation would be leaving something out—namely, the first-personal, experiential, and normative components of meaning that do not lend themselves to the kinds of third-personal explanations that are paradigmatic of the natural sciences). Given the existence of scientific disciplines such as linguistic that *do* have plenty to say about the meaning of language, Macarthur’s objection has, I take it, a much greater claim to plausibility on the weaker reading. As I suggest below, part of the issue here is that Macarthur seems to be operating with a view of science which precludes the possibility that sciences can adopt first-personal methods.

recommended by Price, insofar as it takes itself to be inextricably bound to *scientific* naturalism, will, therefore, be "incapable of explaining conceptual content" (575).

In order to escape this bind, Macarthur thinks that Price needs to embrace a more radical *liberal* naturalism which understands normative phenomena such as reasons, values and meanings as *natural* but non-scientific phenomena. This is a view that Macarthur finds in the works of Robert Brandom, Ludwig Wittgenstein, and John McDowell, among others. Although Macarthur does not explicitly take up the question of whether Price can consistently adopt a broadened conception of naturalism, I believe that the latter can by amending his stance in the following ways.

First, it will require abandoning the rhetorical claim that subject naturalism amounts to "serious science" unlike more orthodox naturalizing projects. For instance, when criticizing object naturalists for tacitly adopting representationalist assumptions which cannot be subject to empirical investigation, Price has a tendency to present subject naturalism as the more properly scientific alternative (Price 2013, 21).[28] But this is ambiguous. It is one thing to suggest that the subject naturalist's refusal to adhere to a theory of thought and language on *a priori* grounds is more consistent with a naturalistic (i.e., empirical) attitude; but it is an entirely different matter to claim that subject naturalists ought to adopt exclusively the methods, explanations and models of the natural (and social) sciences. If Macarthur is correct, then Price has good reason to endorse the former claim but not the later.

Second, embracing a more liberal conception of naturalism will require elaborating on and ultimately expanding subject naturalism's methodological resources. In Chapter Four, I

[28] Macarthur seems to acknowledge that there are passages in Price's writing which suggest that the latter has a broader, more liberal conception of naturalistic explanation in mind (Macarthur 2014, 73 [especially footnote 7]).

shall argue that subject naturalists have good reason to look to philosophical genealogy to supplement their functional explanations of normative phenomena. As it turns out, philosophical genealogy—especially the kind developed by Bernard Williams—is a self-consciously naturalistic approach to philosophical explanation which is not clearly reducible to the kind of third-personal, detached scientific perspective which Macarthur takes to be incapable of properly investigating normative phenomena. In other words, a turn to philosophical genealogy represents a way of moving towards a more liberal form of naturalism.

Although I think that Macarthur is right to push Price's work in the direction of a more liberal naturalism, I wonder whether his own view could use a further radicalization. Macarthur urges that the appropriate contrast between the natural is the supernatural—and not the non-scientific (Macarthur 2014, 74). The problem with this move is that it abandons the aim of articulating the positive relationship between naturalism and science (even if the connection is supposed to be a very loose one). While I share Macarthur's commitment to a broad conception of the sciences which includes the human sciences, I am not entirely persuaded by the strength of his identification of the scientific image with the third-person standpoint. There are, after all, social scientists who explicitly adopt the first-personal perspective through their use of so-called qualitative methods. While Macarthur could deny that these investigations qualify as part of the scientific image, or reject the claim that their content includes normative phenomena, doing so would force him into debates on which he seems reluctant to take a stance.[29] Alternatively Macarthur could alter the criteria of scientificity, downplaying the idea that science is *inherently* third-personal and aimed at producing generalizations (even if just local ones), and instead insist

[29] In particular, Macarthur writes as though he does not want to reinvigorate debates along the lines of the German *methodenstreit.*

on something like a proceduralist approach to science which identifies scientificity with the adoption of a set of shared values—for instance, open-mindedness, fallibilism and responsiveness to criticism. On this picture, liberal naturalism would involve the following equation:

Liberal naturalism = non-supernaturalism + commitment to scientific values.

This would allow for a broad enough conception of nature within which values, meanings and reasons are not rendered mysterious, while preserving a connection—albeit a weaker one—between naturalism and the scientific perspective. Indeed, in an earlier paper, Macarthur comes close to proposing something like this view in his discussion of the scientific method. He writes:

> Following Dewey, we can treat talk of "the scientific method" as shorthand for a loose set of quasi- moral virtues of inquiry together with a fallibilist, experimental attitude to knowledge and understanding. The relevant virtues include acknowledging the need for critical dialogue with others, tolerance of alternative opinions, openness to criticism and the possibility of rethinking one's position on the basis of it - in short, a kind of open-minded democratic ethics of inquiry" (Macarthur 2008, 202).

My suggestion then, is that Macarthur would be better off retaining this conception of science (i.e., as "democratic experimentalism") within his conception of naturalism, while setting aside the idea that scientific inquiry is inherently third-personal. Such a conception, on my view, is well-suited to the purposes of Price's subject naturalism.

Like Macarthur, Paul Redding worries that Price's subject naturalism offers too narrow of a philosophical perspective, ultimately owing to its assimilation of philosophy to the natural sciences. But whereas Macarthur thought that this perspective would give rise to explanatory failures, Redding worries that aligning philosophy too closely with the scientific image, subject naturalism runs the risk of somehow undermining or destroying the basis for many of our normative commitments. In other words, Macarthur can be read as arguing that subject naturalism encounters the problem of completeness, while Redding can be read as arguing that it encounters the problem of transcendence.

To make his argument, Redding situates Price's subject naturalism within analytic philosophy's Kantian legacy, charting a course in which the transcendental aspirations of the latter's critical project have been gradually naturalized, first, from investigating the *a priori* structures of thought, to investigating the structures of language (think *The Tractatus Logico-Philosophicus*), and then second, from this *a priori* concern with language to a post-positivistic picture concerned with contingently-developed (and hence, revisable) linguistic frameworks (think *The Structure of Scientific Revolutions*) (Redding 2010, 264). In so far as it does mark a continuation along the path of naturalized Kantianism, Price's project faces what Redding refers to as the problem of nihilism. This problem concerns the "process of denormativization" which has accompanied the renunciation of philosophy's ambitions to be a foundationalist discipline in migrating towards naturalism. The idea is that naturalism, when pushed too far, engenders not only a kind of epistemological relativism, but also a loss of the first-person, engaged perspective from which to adequately make sense of our normative commitments (266-7). Redding takes this to be succinctly captured in Bernard Williams claim that "reflection can destroy knowledge" (267). As Williams himself elaborated this concern, the detached, explanatory perspective central to philosophical naturalism risks abandoning the kind of stance from which they might be able to make sense of our commitments and values (274). For Williams, of course, this argument does not entail a rejection of naturalism, but rather a frank recognition of its limits and risks.

These are implications that Price should by his own lights be eager to avoid. After all, he not only wants subject naturalism to be able to *explain* normative phenomena, but at least minimally, to be able to support or sustain people's commitments to those phenomena. For instance, as we shall see below, Price's account of assertion is not simply meant to *explain* how or why a central dimension of our discursive lives takes the shape that it does. Rather, he is, in some sense, interested in offering a kind of justification (or to borrow a term from Bernard Williams) a *vindication* of the normative dimension of truth which he discerns (Price 2003).[30] This desire is perhaps most evident in Price's strong resistance to Richard Rorty's claim that there is no interesting difference between justification and truth (Rorty, 1998; Price 2003). On what basis is the global expressivist's normative ambitions about truth supposed to be understood? It is one thing to engage in what Price calls "linguistic anthropology" or "linguistic genealogy" with the aim of *explaining* the function of certain aspects of linguistic usage. But it is quite another task to motivate or justify, within a naturalistic perspective, normative claims about those practices—whether it be regarding their indispensability for a "meaningful human life" or to recommend "correctives" to usage that has been, in some sense, corrupted by philosophical impositions (Price 1988, 174).

Redding's discussion, however, is surely underdeveloped. In particular, he offers little to no explanation of *why* or *how* exactly the kind of detached, reflective standpoint afforded by the scientific perspective is supposed to *undermine* (as opposed to simply fail to explain) norms, values or commitments. Here he seems to rely on Bernard Williams' discussion of these issues in *Ethics and the Limits of Philosophy*, which is a helpful starting point, but by no means an uncontroversial one. What is it, exactly, about naturalistic (here read as broadly scientific)

[30] In chapter 8 of *Facts and the Function of Truth* Price briefly discusses the question of whether the kind of detached perspective from which his explanatory account of truth is advanced is capable of "prescribing" linguistic usage (Price 1988, 174). I discuss his view and its connection with Bernard Williams' later in this chapter. For Williams' discussion of vindicatory naturalistic explanations (in the form of genealogies) see Williams (2002, 36).

perspectives on our own linguistic practices, that is supposed to threaten practitioners' commitments held within those practices?

In Chapter Four I shall explore this potential problem of normative commitment in more detail. Following Redding's lead I shall take up the work of Bernard Williams in order to assess whether or not the global expressivist might be able to find her normative footing. Ultimately, I shall conclude that there are reasons to believe that Price's form of pragmatism requires methodological amendments, and that one promising way of making them is to turn to a more fully-elaborated form of philosophical genealogy. It is the task of explicating and developing these amendments that the dissertation on a whole pursues. But setting aside these considerations for the moment, I now turn to expounding the positive dimension of Price's global expressivism.

2.3 Global Expressivism: Pragmatism and Metaphysical Quietism

Increasingly, Price has come to conceive of his positive project as a kind of pragmatism, which he characterizes in terms of two key components. First, a non-representationalist view of thought and language, which, as we have seen, involves setting aside substantive semantic notions for theoretical purposes. Second, for Price, pragmatism affords "linguistic priority" to the kinds of questions that tend to generate placement problems. By resisting (what was described above as) the material approach to philosophical problems, pragmatists are better able to keep their theoretical vision focused on questions about what human beings *do* with language (Macarthur and Price 2007, 232-3). Rather than ask, for example, what goodness, probability, or necessity *is,* the pragmatist thinks that we are better off asking how it is that humans came to employ those concepts, the idea being that once that question is answered there will simply be no need to raise the more metaphysical-sounding questions. Pragmatists, for Price, are metaphysical quietists.

But what does it mean to be a metaphysical quietist, exactly? Surely having ontological commitments is *unavoidable* for any naturalistic inquiry. After all, by Price's own lights he is committed to, at the very least, the existence of human beings, natural environments, language, and so on. It is important to note that Price is not denying any of this. In fact, he is quite explicit about his willingness to "stand with the folk, and to affirm the first-order truths of the domain in question—to affirm that there are beliefs, and values, and causes, and ways things might have been" (Macarthur and Price 2007, 236).[31] Rather, what the pragmatist *rejects* "is any metaphysical theoretical perspective from which to say more about these matters—that they do or don't *really* exist, that they are *really* something subjective, or whatever" (236).[32] What is the problem with such a perspective? Here Price takes a page out of Rudolf Carnap's anti-metaphysical playbook—invoking the latter's distinction between "internal" and "external" questions.

For Carnap, ontological questions are legitimate only in so far as they are asked "internally" to what he called a "linguistic framework." A linguistic framework consists of terms, along with a set of rules for employing them, examples of which included the frameworks

[31] Price sometimes puts the contrast between his own deflationary, expressivist approach and the various approaches to placement problems encouraged by representationalism in terms of a distinction between ontologically conservative and ontologically non-conservative theories (Price 2009b, 262-30). Non-representationalist semantic views are ontologically *conservative* insofar as they set aside questions about *referents* and truthmakers. Yet, it would be misleading to read Price as identifying metaphysics with simply holding ontological commitments. The point rather, as Michael Williams nicely puts it, is that on the kind of subject naturalist approach recommended by Price, the only "antecedent ontological commitments… are to speakers, their utterances and so on: that is, to things that everyone is bound to recognize anyway" (Williams 2013, 130).

[32] One way of putting this would be to say that for Price, ontological commitments are always 'internal' to some inquiry or discursive practice.

for talking about “the world of things,”[33] or the “system of numbers” (Carnap 1956, 208). It is appropriate to raise ontological questions about, for instance, *tables* or *numbers* within such frameworks because the latter carry with them empirical and logical methods for answering such questions. It is when ontological questions are raised “externally” to linguistic frameworks, for instance when asking *about* the existence or reality of the frameworks themselves, that philosophical confusions are liable to emerge (214). This is because the investigator advancing such questions has taken up a position from which the rules or procedures required to answer them are unavailable. This does not mean that *all* external questions are senseless. It is always possible to raise “pragmatic” questions about a framework, such as whether it is “more or less expedient, fruitful, [or] conducive to the aim for which the language is intended” (214). But, as Price points out, raising an external question about a linguistic framework will involve mentioning rather than using the resources of that framework (Price 2009a, 284). Thus, philosophers who tend to go in for metaphysical (as opposed to pragmatic) “external” questions have thereby committed something like a category mistake.[34]

In adopting this Carnapian distinction, Price insists that he is *not* thereby committed to a metaphysical view (at least in any standard sense of the term ‘metaphysics’). That is, Price is neither committing himself to something like an anti-realist position (construed as a theoretical claim that there are *not* certain kinds of entities or facts), nor is he committed to a kind of metaphysical pluralism (construed as the claim that what there *really* is, is whatever our multifarious linguistic frameworks commit us to). These claims presuppose precisely the kind of

[33] That is, “the spatio-temporally ordered system of observable things and events” (Carnap 206-7).

[34] Price writes that, “Legitimate *uses* of terms such as “number” and “material object” are necessarily internal, for it is conformity (more-or-less) to the rules of the framework in question that constitutes use. But as internal questions, as Carnap notes, these questions could not have the significance that traditional metaphysics takes them to have. Metaphysics tries to locate them somewhere else, but thereby commits a use-mention fallacy. The only legitimate external questions simply *mention* the terms in question” (Price 2009a, 283).

"external" perspective that Price finds unhospitable to any kind of ontological claim. But how is he to avoid making them? It is here that Price invokes a distinction between *active* and *passive* rejection of a thesis. Rather than "*denying in one's theoretical voice*" that the expressions within some linguistic framework *genuinely refer* or *have truth conditions*, he argues that it is possible to "*remain silent in one's theoretical voice*" about such questions (Price 2013, 12).[35]

[35] A number of recent commentators have taken issue with Price's anti-metaphysical ambitions. There are, broadly speaking, two different directions from which this charge tends to be made. On the one hand, some philosophers deny that it is either possible or desirable to avoid metaphysics (Blackburn 2013; Lynch 2015; Legg and Giladi 2018). From this perspective, Price's project appears at best naïve (in the sense that it ignores pressing issues) or at worst self-deceptive (in so far as it remains uncritically ignorant of its own inevitable metaphysical commitments). On the other hand, there are those who are sympathetic to Price's desire to avoid metaphysics, but find his strategies for doing so to be lacking (Horwich 2013; Knowles 2017).

Catherine Legg and Paul Giladi have positioned themselves in this first camp, arguing that global expressivism inevitably involves a number of metaphysical commitments which threaten to render Price's "anti-metaphysical" attitude inconsistent. For instance, these authors claim to find a number of "dualisms" underpinning Price's entire project. These include his distinctions between "i-representation" and "e-representation," between the material and linguistic starting points of placement problems, as well as the distinction between the space of reasons and the space of causes (Legg and Giladi 2018, 70-72). Not only are these taken to be "fundamentally metaphysical" in the sense that they are "concerned with a sufficiently general inquiry into the nature and structure of reality" (71), but, perhaps more seriously, because such commitments take the forms of dualisms they are the kinds of commitments that a pragmatist should reject (71). In a similar vein, Legg and Giladi see Price's distinction between object and subject naturalism to be problematic because it entails a view of the human subject as "divorced from its broader context of surrounding objects"—which, again, strikes them not only as a metaphysical thesis, but an implausible one at that (75).

Moreover, even if one were to overlook these apparent dualisms, Legg and Giladi find Price's metaphysical quietism to be obviously self-refuting. After all, any naturalist is surely committed to a picture of human beings as "normative, self-reflecting discursive agents" (Legg and Giladi 2018, 73). But, they contend, this is clearly to engage in metaphysics. After all, "by conceiving of ourselves in this way, he is inevitably engaging in some kind of general inquiry into the nature and structure of reality, as humans are (of course) themselves part of reality" (73). As this passage might suggests, these critics are troubled by the fact that Price seems to take too narrow a view of metaphyics. As they put it,

> "for Price, metaphysics consists in the holding of ontological commitments understood solely as specific objects in the world which 'hang off' our true sentences and (to use his picture-book metaphor) have the same 'shape', somehow. Metaphysics can involve much more than this, and has done so in a rich tradition stretching back 2000 years. We think it a pity that Price did not consider a few alternative conceptions of the discipline before dismissing it so wholeheartedly" (77).

I think that these authors are correct to note that Price (like Rorty, whose views they associate with him) is, indeed, committed to a picture of human beings as normative, self-reflective, agents. Moreover, as we have already seen, Price can wholeheartedly accept that the sciences (and presumably any linguistic practice) will inevitably carry with it ontological commitments, for instance, that there are human beings, who inhabit an environment, and so on. But as Rydenfelt points out, rightly on my view, Legg and Giladi's charge that this contradicts Price's "metaphysical quietism" only has bite if one accepts the "dubious claim" that for Price having a metaphysics is coextensive with having some ontological commitments (Rydenfelt 2019). A more charitable construal of Price's project would be to view it as recommending a quietism about the kinds of philosophical projects which attempt to inquire into the reality of those commitments by asking something like a Carnapian "external question." So, while I think that these

This might seem like a radical view. But Price is quick to maintain that his intentions are less "iconoclastic" than some of his anti-representationalist forebearers (Price 2013, 26). While willing to endorse metaphysical quietism about certain topics, he is eager to avoid a more general philosophical quietism, which denies that there is anything of philosophical interest to be said of such topics as meaning, mind, morality, etc., (Macarthur and Price 2007, 236).[36] According to Price, philosophical naturalists can still make valuable contributions to such topics, he just thinks that the most influential attempts to do so have held assumptions, which we would be better off relinquishing. Hence, the need for a reconstructed philosophical naturalism.

It turns out that many of the resources required to pursue this philosophical position are already in play. To this end, Price conceives of his own role as a kind of "philosophical real estate agent," (Price 2011, 5) or as a kind of "trail marker" as opposed to a "trail blazer" (Price 2013, 26). As we shall see, these self-ascriptions have an air of false modesty, because he is also

authors underestimate the resources that Price has at his disposal to answer to their concerns, their article is helpful insofar as it represents a common position that a global expressivist is likely to encounter. Moreover, they point to an issue on which Price could have been clearer: what exactly does the global expressivist mean by 'metaphysics'?

On the opposite end of the "metaphysical spectrum," so to speak, Jonathan Knowles argues that Price's current formulation of his position does not do enough to rule out metaphysics. In particular, Knowles takes it that one of Price's ambitions is to show that the kinds of metaphysical projects manifest in contemporary approaches to placement problems are in some sense "irrational" or "unmotivated," and that this is more or less accomplished by showing that object naturalism is ruled out once the case for non-representationalism is made (Knowles 2017, 4783). But, Knowles argues, this does not follow. As he explains, "It seems what Price says at most suggests that if we adopt a material starting point for placement problems, we owe some account of why we should prosecute them in a metaphysical way, rather than go in for [global expressivism]. But this falls short of showing that this project is incoherent or somehow irrational" (4786). To press this point, he sketches the outline of a non-repesentationalist position which accepts naturalism, but for which it makes perfectly good sense to go in for a kind of version of the sorts of placement problems constitutive of object naturalism. Such a position (which Knowles associates with Quine) would insist on a kind of 'reduction' of different vocabularies to scientific ones, but in a way that did not depend on substantive semantic terms (4790-1). These considerations, moreover, would seem to threaten Price's reliance on the Carnapian internal-external distinction to secure the anti-metaphysical conclusion that he wants. Although he does not ultimately develop it in his paper, Knowles suggests that in order to avoid these unacceptable implications, the global expressivist requires a commitment to a kind of radical anti-reductionism (4795).

[36] As Price and Macarthur explain, "[o]ur pragmatists are metaphysical quietists. But note that they are not philosophical quietists *tout court*, of there could be such a view. On the contrary, they take some relevant theoretical matters very seriously indeed: in particular, some broadly anthropological issues about the roles and genealogy of various aspects of human linguistic behavior" (236).

an adept practitioner of linguistic anthropology. One promising path towards global expressivism emerges once the effort is made to bridge a gap between two influential research projects. The first, exemplified by Simon Blackburn, is a sophisticated form of expressivism called quasi-realism. The second is Robert Brandom's semantic inferentialism. Taken together, Price suggests, these two views—at least when suitably amended—point the way to "the pragmatist promised land" (Price 2013, 32).

2.3.1 From Local to Global Expressivism: Quasi-Realism and the Bifurcation Thesis

Throughout his career, Price has found an important interlocutor and philosophical ally in Simon Blackburn.[37] Blackburn's quasi-realism begins with the expressivist insight that, contrary to grammatical or syntactical appearances, some areas of discourse do not function primarily to describe or accurately represent reality, but are better understood as performing some other function—such as expressing attitudes or commitments. Although best-known as a metaethical position, expressivism presents a compelling way of treating a variety of philosophical topics—including probability, causation, conditionals, and logic, among others. For Price, quasi-realism marks a step in exactly the right direction, exemplifying non-representationalist, subject naturalist orientation which remains sensitive to the different functions that language serves.

Unlike its emotivist predecessors, quasi-realism tries to capture many of the attractions of philosophical realism, but in a less metaphysically committed way. The quasi-realist aims to tell a naturalistic story about their target discourse, such that: (i) it is understood primarily non-representationally (for instance, that it as expressive of our attitudes, sentiments and

[37] See especially (Price, 1988; Price 1992; Price, 1993; Macarthur and Price 2007) as well as Price's reviews of some of Blackburn's books (Price, 1996; and Price, 2006).

commitments);[38] and (ii), on the basis of which we come to understand *why* that domain of discourse takes the shape it does, such that the quasi-realist "earns the right" to construe its claims as being true or false, reasonable or unreasonable.

In his book *Ruling Passions* Blackburn presents his most detailed version of quasi-realism as a specifically metaethical theory. The claim is that the expressivist can capture the scope and complexity of human moral practices, in which duties, virtues and reasoning about ends are situated within a naturalized theory of normativity and practical reasoning. In addition to its ability to explain why ethical discourse has all the "trappings" of regular descriptive discourse, Blackburn's quasi-realism departs in other respects, from earlier logical-empiricist versions of expressivism. First, in its adoption of a holistic conception of mind and meaning (Blackburn 1998, 51-9). And second, in taking advantage of the resources of a deflationary or minimalist conception of truth (77-83).[39]

Despite his sympathies, Price thinks that most forms of expressivism (including Blackburn's) suffer from a serious flaw. Expressivists typically want to contrast their non-representational account of some area of language with some other genuinely representational

[38] In the metaethical cases, for instance, the range of attitudes that he has in mind is meant to be as exhaustive as possible. Blackburn proposes the notion of a "staircase of practical and emotional ascent" (Blackburn 1998, 9). At the "lowest level" of the staircase we find "simple preferences", like tastes and aversions. At another level we find "reactions to reactions" which consists in preferences or attitudes that we would prefer that others share (e.g., if I feel that you are unjustified at being angry with someone else, I might express my disagreement with your sentiment by saying that it is none of your business (9)). As we continue higher on the staircase, we find attitudes that we regard as *compulsory* such that we "become prepared to express hostility to those who do not themselves share [them]" (9). For Blackburn, ethics begins with and trades in these "higher level" attitudes, and he follows Hume and Smith in understanding them as stemming from the fact that human beings are social creatures fundamentally disposed to take up the common point of view (1998, chapter 7).

[39] For Blackburn, a holistic view of the mind means that "a person's entire mentality forms a kind of web or field or force in which no single element has its own self-standing connection with action. Different beliefs and desires (and perhaps other states, such as emotions, attitudes, wishes, fantasies, fears, and of course values) come together to issue in action. But the contribution of any one of them will vary according to what else is in the mix, and therefore resists definition in terms of behavior" (Blackburn 1998, 52).

area (e.g., the claims made by natural scientists).[40] Price refers to this as the *bifurcation thesis*: the idea that one can draw a sharp line within language between those genuinely representational and non-representational uses (Price 2013, 30).[41] The most direct attack on this idea can be found in "Pragmatism, Quasi-Realism, and the Global Challenge," where Price and David Macarthur advance two lines of argument against the tenability of merely local forms of expressivism. The first mobilizes and ultimately aims to reverse an external criticism often directed at expressivism. The second tries to show that unless expressivists abandon the bifurcation thesis, they will become victims of their own success. Allow me to consider these in turn.

A well-known argument against expressivism, which James Dreier calls *the problem of creeping minimalism*, aims to exploit its apparent incompatibility with semantic deflationism (Dreier 2004). Typically, the expressivist wants to advance some version of the following claims:

[C1] Some areas of discourse are (genuinely) cognitive.

[C2] Some target area of discourse, T (e.g., moral discourse) is non-cognitive.

[C3] T performs some (non-cognitive) function F.

[40] Commenting on this tendency, Michael Williams writes: "That the expressive function of a particular vocabulary item explains its assertional and inferential use properties, themselves specifiable in an ontologically conservative way, is the local expressivist's deep insight. The tendency to take this insight to imply that the vocabulary to which his analysis applies is not 'really' descriptive is his *ur*-mistake" (Williams 2013, 238).

[41] One of Price's longstanding projects has been to show that such a distinction is unmotivated. Indeed, the first part of *Facts and the Function of Truth* develops a "skeptical strategy" aimed to show that there can be no principled way of demarcating fact-stating discourse from non-fact-stating discourse through an analysis of statementhood (Price 1988).

But, the argument goes, if one accepts semantic deflationism, then *any* sentences that are both disciplined and syntactically well-formed will be trivially cognitive. Therefore, C2 will be trivially false (Macarthur and Price 2007, 240).[42]

Price and Macarthur find this argument to be misguided, and argue that deflationism actually lends support for expressivism—that is, so long as the expressivist is willing to adopt a more thoroughgoing nonrepresentationalism (i.e., global expressivism). The traditional expressivist ultimately wants to make *both* a set of negative and positive claims (Macarthur and Price 2007, 240). On the one hand, they want to deny that certain terms or statements possess substantive semantic features. This is captured by C2 above, along with its (often tacit) contrast with C1. On the other hand, the expressivist advances some positive, *non-semantic* alternative account of the linguistic function in question. This is captured by C3 above. Where the argument from creeping minimalism misfires, is in assuming that accepting semantic deflationism requires abandoning C2 *in a way that entails cognitivism* (i.e., that as "theoreticians, we must endorse its negation" [240]). One reason that this entailment might seem inevitable is that the expressivist's negative claim appears to be "a substantive theoretical claim, cast (essentially) in semantic terms" (240).[43] But for Price and Macarthur, to accept this would be to overlook a subtle option available to the expressivist. In the same way that someone who rejects theological debates altogether might wish to abstain from making any substantive claim about the nonexistence of God, it is open to the expressivist to allow their negative claim to be deflated, thereby refraining

[42]Moreover, the result is that expressivism now looks indistinguishable from realism. As Dreier puts it, "Minimalism sucks the substance out of heavy-duty metaphysical concepts. If successful, it can help Expressivism recapture the ordinary realist language of ethics. But in so doing it also threatens to make irrealism indistinguishable from realism. That is the problem of Creeping Minimalism" (Dreier 2004, 26). This reason has led many expressivists to resist the label of 'non-cognitivism' (See Copp and Blackburn, 2005).

[43] I take this point to be directly related to Price's argument, discussed above, according to which the object naturalist is incapable of taking a contingent attitude towards the question of whether semantic terms themselves stand in semantic relations or have semantic properties.

from having to endorse in their “theoretic voice” its negation (i.e., cognitivism) (240-1). This, as we have already seen, is the difference between *active* and *passive* denial.

Because the expressivist’s negative and positive claims are independent of one another, endorsing semantic deflationism leaves open the option to *simply say nothing of theoretical significance about whether the discourse in question is genuinely representational or cognitive*, while effectively leaving the positive (explanatory) thesis untouched (Macarthur and Price 2007, 241). That is, an expressivist’s positive account supporting some instance of C3 can still be consistent with semantic deflationism so long as she does not actively deny C2. However, as I mentioned above, expressivists typically want to *contrast* their target domain with some other areas of discourse that *are* in some sense, genuinely representational, cognitive, or truth-apt: hence their commitment to something like C1. What the above line of reasoning is supposed to show, however, is that this attempt at invoking a bifurcation between genuine and, say, “quasi” truth-evaluable discourses is ruled out by deflationism.[44] The result is, according to Price, an argument in favor of a global version of expressivism.

It would seem then, that the local expressivist faces a dilemma: they must either abandon semantic deflationism (and attempt to argue for the bifurcation thesis on some principled basis); or they must give up the bifurcation thesis, thereby embracing a global expressivism. Here Price and Macarthur point to factors internal to expressivism—at least internal to more sophisticated versions like Blackburn’s—that motivate the latter option. Above, I mentioned that a theoretical advance of quasi-realism over its emotivist predecessors lies in its attempt to account for the

[44] As Price and Macarthur explain, “Deflationism disallows this question [i.e., “is some domain genuinely representational?”], and thereby the contrast that depends on it—but it doesn’t disallow the expressivist’s positive, pragmatic account of what supposedly lies on the non-representational side of the fence” (Macarthur and Price 2007, 241).

descriptive or cognitive “shape” of moral discourse. This virtue, the claim goes, ought to push the local expressivist to go global. For, as Price and Macarthur put it, suppose that the quasi-realist

> Succeed(s) in explaining, on expressivist foundations, why non-descriptive claims behave like… genuine descriptive claims. If these explanations work in the hard cases, such as moral and aesthetic judgments, then it seems likely that they’ll work in the easy cases, too—i.e., for scientific judgements (Macarthur and Price 2007, 245).

But if this is true, the quasi-realist is in a rather uncomfortable position. On the one hand, they owe an explanation of what work the notion of “genuine descriptive claims” is supposed to be doing (the global expressivist’s suspicion, of course, is that such an explanation is unlikely to be found). On the other hand, if it turns out that there *is* some work for such a class of claims to be doing, then it would seem that the quasi-realist has not succeeded in their original task of “explaining how non-descriptive discourse can emulate the real thing” (245). Again, Price and Macarthur think that the local expressivist’s best option is clear: abandon the last vestiges of representationalism by setting aside the bifurcation thesis.

2.3.2 Pragmatism in Two ‘Tiers’

But what exactly does it mean for the local expressivist to ‘go global’? For Price, the key is to appreciate the possibility of adopting a two-tier pragmatist explanatory framework (Price 2013, 153). On the ‘lower level,’ the pragmatist can continue to take up the kind of explanatory stance exemplified by traditional local expressivists. That is, they can advance naturalistic stories about the various functions that our diverse discursive practices serve. That we talk, say, of *values*, *probabilities*, *mental phenomena*, and so on, is to be explained by virtue of the fact that such types of commitments and talk “have different origins in our complex natures and relations to our physical and social environments” (Price 2013, 33). The goal here is to account for a kind of ‘functional pluralism’ within language, which Price associates with Wittgenstein.[45] Attention to this kind of plurality is, in an important respect, the starting point for the expressivist’s insistence that *some* language games are (despite appearances) not primarily in the business of describing how things are. As Price puts it, the explanatory common denominator of such approaches is that:

> particular, contingent features of a creature’s practical circumstances—e.g., that she is a *decision-make under uncertainty*, or an *agent*, or a bearer of *epistemic dispositions*—provide the source of variability… Each of these features constitutes what we might call a *practical stance*—a practical situation or characteristic that creature must instantiate if the language game in question is to play a defining role in her life (48).

As we have already seen, these local forms of expressivism typically end up invoking a contrast between non-representational (or perhaps quasi-representational) vocabularies with *genuinely* representational ones, thus invoking some form of the bifurcation thesis. Price clearly thinks that this is not a desirable option. Hence, his suggestion that the global expressivist requires an additional ‘upper level’ approach to assertoric practices *in general* which satisfies two conditions. First, it must be non-representational (i.e., it cannot make use of substantive semantic word-world relations). And second, it must apply in a uniform way across the various local cases. Fortunately for Price, not only has such an approach been worked out in detail, but it turns out that there are options to choose from. On the one hand, the global expressivist can look to the account of assertion found in Robert Brandom’s semantic inferentialism. On the other hand, Price has worked out his own account of truth and assertion that is, supposedly, just as well suited for the job.[46] I shall briefly discuss these two positions in turn.

[45] See especially (Price 2010).

[46] One question about which Price could be clearer concerns the extent to which Brandom’s or his own account are compatible. Price has criticized Brandom for conceding too much to the representationalist (Price 2010). Brandom, in turn, has voiced his own suspicions about Price’s project—in particular, that it doesn’t do justice to the important

Brandom's account of assertion is, if not the backbone, then at least an important vertebra of his more general theory of meaning, called *semantic inferentialism*. The inferentialist's fundamental insight is that one can explain the meanings of linguistic expressions (including speech and thought) on the basis of the role that they play in inference, or reasoning. Following Wilfrid Sellars, Brandom's point of departure is to account for the meanings of expressions by attending to their role within the social, norm-governed linguistic practices of what he calls 'the game of giving and asking for reasons.' This "explanatory strategy," as he puts it, involves two key steps. First, it is to:

> begin with an account of social practices, identify the particular structure they must exhibit in order to qualify as specifically *linguistic* practices, and then [second] to consider what different sorts of semantic contents those practices can confer on states, performances, and expressions caught up in them in suitable ways (Brandom 1994, xiii).

The first step involves giving an account of *normative pragmatics,* which aims to understand the norms governing language-users as arising out of and operating within social practice.[47] The second step involves giving an account of *inferentialist semantics,* according to which, linguistic expressions acquire semantic content on the basis of their role within this inferential practice. [48] Both aspects of Brandom's account build on one of Wilfrid Sellars' central insights: that grasping a concept is not merely a matter of being able to reliably respond to one's environment,

expressive role that representational vocabulary plays within discursive practice (Brandom 2013, 109). In particular, Brandom doesn't think that Price's notion of i-representation (explicated below) actually *gives* us an account of *representation* (but rather, something like semantic content) (Brandom 2013, 106).

[47] In particular, Brandom is ultimately interested in explicating *assertion* in terms of its place within norm-governed game of giving and asking for reasons.

[48] More specifically, Brandom's strategy is to begin his account of semantics with the notion of propositional contentfulness, and to explicate propositional contentfulness in terms of "the *inferential* articulation of the social practice of giving and asking for reasons" (Brandom 1994, 79).

but requires the practical ability to correctly draw inferences in which that concept is involved (Brandom 2000a, 48).[49]

On the one hand, Brandom gleans from this insight an answer to the question of what it is for a social practice to qualify as a *linguistic* practice. Plausibly, its participants must be able wield *concepts*, which for Brandom means having "practical mastery over the inferences [they are] involved in" (2000a, 48). This practical ability, of course, requires that language users take themselves and others to be subject to a range of normative constraints. As he puts it,

> Saying or thinking *that* things are thus-and-so is undertaking a distinctively *inferentially* articulated commitment: putting it forward as a fit premise for further premises, that is, *authorizing* its use as such a premise, and undertaking *responsibility* to entitle oneself to that commitment, to vindicate one's authority, under suitable circumstances, paradigmatically by exhibiting it as the conclusion of an inference from other such commitments to which one is or can become entitled (Brandom 2000a, 11).

This, in turn, requires not only that speakers are able to make commitments and attribute them to others, but that they are able to keep track of their own and each other's commitments and entitlements.[50] A key device for Brandom is to construe discursive practice in terms of "deontic

[49] This argument is famously developed in Sellars (1956/1997). A parrot or a thermostat, when appropriately trained or properly functioning, can possess the former ability; but presumably not the latter. That is, a parrot could be taught to squawk 'that's red' whenever it is presented with red objects; but, to the extent that it lacked the ability to draw inferences, such as 'that's red, therefore it is not blue' is the extent to which it lacks *concepts*. As Brandom puts it, "to grasp or understand a concept is to have practical mastery over the inferences it is involved in" (2000a, 48). Moreover, for Brandom, the ability to wield concepts marks the boundaries between *sentience* (understood as "as the capacity to be *aware* in the sense of being *awake*" [(Brandom 1994, 5), see also (2000a, 2; 157)] and *sapience* (understood in terms of "understanding or intelligence"). To treat something as sapient means attributing to it "belief and desire as constituting reasons for their behavior" (Brandom 1994, 5).

[50] As Brandom explains: "Specifically linguistic practices are those in which some performances are accorded the significance of assertions or claimings—the undertaking of inferentially articulated (and so propositionally contentful) commitments. Mastering such linguistic practices is a matter of learning how to keep score on the inferentially articulated commitments and entitlements of various interlocutors, oneself included. Understanding a speech act—grasping its discursive significance—is being able to attribute the right commitments in response. This is knowing how to change the score of what the performer and the audience are committed and entitled to (Brandom 2000a, 164-5).

scorekeeping," according to which "the significance of a speech act is how it changes what commitments and entitlements one attributes and acknowledges" (Brandom 2000a, 81).

On the other hand, Brandom's second step is to explain how linguistic expressions—the fundamental 'moves' in this norm-governed game—come to acquire meaning through their ability to serve as premises and conclusions in material inferences. Again, he follows Sellars in explicating conceptual content in terms of inferential role (Brandom 1994, 89). In this respect, Brandom invokes a pragmatist (as opposed to a 'Platonist') order of semantic explanation, which is to say that it aims to explain *meaning* in terms of *use*, and not the other way around (Brandom 2000a, 4). As he puts it, this "conceptual pragmatism":

> Offers an account of knowing (or believing, or saying) *that* such and such is the case in terms of knowing *how* (being able) to *do* something. It approaches the contents of conceptually explicit propositions or principles from the direction of what is *implicit* in practices of using expressions and acquiring and deploying beliefs (Brandom 2000a, 4).

From the perspective of Price's global expressivism, Brandom's project is especially compelling because it clearly satisfies the two desiderata mentioned above. First, Brandom's account of meaning is avowedly *non-representationalist*.[51] On the one hand, this follows from his methodological commitment to conceptual pragmatism: by beginning with an account of normative pragmatics he refuses to afford substantive semantic concepts any *explanatory* role in

[51] Brandom rejects representationalism in its traditional, Cartesian form whereby "Awareness was understood in representational terms—whether taking the form of direct awareness of representings or of indirect awareness of represented via representations of them" (7). That being said, Brandom is still interested in offering an *expressivist* account of representation (2000a 10; and esp. chapter 5). This ambition frequently puts him at odds with pragmatists like Rorty who want to do away with the notion of representation entirely.

his theory.[52] On the other hand, Brandom has worked out and actively defended his own account of semantic deflationism.[53]

The second reason that Brandom's approach is especially well-suited for the purposes of Price's "global" level of expressivism, is that it affords special status to *assertions* and *propositional content* as both the fundamental speech act and unit of meaning. Briefly: the inferentialist ultimately wants to explain conceptual content in terms of the *use* of linguistic expressions within inferential practice; but since the latter trades in propositions, this "entails treating the sort of conceptual content that is expressed by whole declarative sentence as prior in the order of semantic explanation to the sort of content that is expressed by subsentential expressions such as singular terms and predicates" (Brandom 2000a, 12-13). This is especially important because, on this picture, there are no differences in "kind" between different semantic contents—they are all conferred in the same way. As a result, the inferentialist jettisons any stance from which to make sense of the local expressivist's contrast between "genuine" or "merely quasi" assertions. Contrasting this approach to meaning with Wittgenstein's famous metaphor of language as a heterogeneous *city*, Brandom writes:

> [T]he inferential identification of the conceptual claims that language (discursive practice) has a *center*; it is not a motley. Inferential practices of producing and consuming *reasons* are *downtown* in the region of linguistic practice. Suburban linguistic practices

[52] That is, Brandom's account of normative pragmatics explicitly avoids presupposing semantic terms like *meaning, reference*, *truth* and the like, in offering an account of assertion in terms of the game of giving and asking for reasons.

[53] Brandom has developed a *prosentential* account of truth. The *inspiration* for this deflationary approach comes from the work of Dorothy Grover, Joseph Camp and Nuel Belnap. Its basic idea is to treat the truth predicate (i.e., expressions of the form 'x is true') as a logical operator for creating a "prosentence"—conceived along the lines of a pronoun. Just as pronouns like 'she' allow speakers to the effectively pick out the referent of previously-employed nouns, so too prosentences enable speakers to refer to previously uttered statements. For example, in the expression "Judy likes to sing; *she* is also a talented dancer", the word "*she"* enables the speaker to refer to Judy. Likewise, suppose someone were to claim, "Judy is a talented dancer." Then, another speaker might respond with "that's true", thereby making use of the previously uttered sentence. For a detailed discussion see (Brandom 2009, chapter 6; Brandom 1994, chapter 5; Tebben, 2015).

> utilize and depend on the conceptual contents forged in the game of giving and asking for reasons are parasitic on it (Brandom 2000a 14-5).

By appealing to the inferentialist's account of assertion, Price thinks that the global expressivist is equipped with an important tool to provide a naturalistic account of thought and language. Not only does such an account eschew representationalism, but it can make sense of the functional diversity of language.[54] On the one hand, "lower level" inquiries into the various functions and genealogies of our diverse language games promise to deliver (subject) naturalistic explanations of human linguistic behavior that leave no room for the kinds of metaphysical worries typically manifested in placement problems. On the other hand, the "upper level" pragmatic account of assertion prevents the slip back to local expressivism by disallowing the pragmatist to draw a meaningful distinction between genuine and "merely quasi" assertions.

While Price is hopeful that Brandom's semantic inferentialism can be appropriated for the purposes of global expressivism, it is important to note that Price's own longstanding account of truth and assertion is, ostensibly, capable of performing the same job. As he puts it, "at its simplest, my proposal is that the assertoric language game is simply a coordination device for social creatures, whose welfare depends on collaborative action" (Price 2013, 49). Since my critical discussion of Price offered below will focus on this aspect of Price's theory, it is worth discussing it somewhat in detail.

One of Price's earliest and most comprehensive discussions of truth can be found in the second part of his 1988 *Facts and the Function of Truth*. There, he urges philosophers to abandon hope for an *analysis* of truth, suggesting instead that they aim to explain the role that the

[54] At least when it comes to the project of defending philosophical naturalism.

concept plays in human life (Price 1988, 119). His own account takes the form of a functional-genetic explanation of how truth-talk may have arisen. The guiding idea is that truth and falsity serve to encourage a useful kind of linguistic behavior—namely reasoned argument—whose long-term advantage consists in the fact that it encourages speakers to form their beliefs by pooling cognitive resources. That is, by deploying the terms 'true' and 'false' as part of a normative system of punishment and reward, our early hominid ancestors' disagreements were such as to generate social instability which could only be resolved through argument or dialogue. These forms of linguistic behavior:

> ensure that as individuals we hold and act on attitudes that reflect, to some extent the combined wisdom of our linguistic community. Our behavioural dispositions can thus be tested against those of other speakers, before they are put to use in the world. The guiding principle is that it is better to be criticized for claiming that tigers are harmless than to discover one's mistake in the flesh (Price 1988, 145).

Whatever its merits, Price's hypothesis that the normative character of truth may have emerged in light of evolutionary considerations does little *by itself* to show that we actually find such a norm in practice, or that it is as widespread as Price thinks. Perhaps more importantly, such an argument—on pains of committing the genetic fallacy—does even less to motivate the claim that such a norm *should* play a role in current practices.

Fortunately, Price has an independent argument, intended to highlight both the ubiquity and indispensability of the normative character of truth. His strategy is to identify three norms of assertion—roughly, *sincerity, justification,* and *truth*—and then to imagine a community of language users who lacked the norm of truth. This approach is meant to show just how central the norm is to human life. To see how this is supposed to work, consider the first norm of sincerity (or subjective assertibility) which says that "a speaker is incorrect to assert that P if she does not believe that P" (Price 2003, 168). That this norm differs from truth is evinced by its applicability to a range of both indicative and non-indicative utterances. In much the same way that I open myself up to charges of insincerity for *claiming* things which I do not believe, it is (typically) inappropriate for me to *request* something that I do not actually want.

The second norm—of justification, or warranted assertability—says that "a speaker is incorrect to assert that P if she does not have adequate (personal) grounds for believing that P" (Price 2003, 169). As the parentheses indicate, this norm comes in two flavors. Its weaker form indicates that when a speaker asserts something for which she lacks adequate personal evidence, others are justified in disapproving of her. In its stronger sense, this norm expands the relevant sense of warrant to the level of a speaker's community. In either case, Price argues that sincerity and justification are insufficient to account for our practices of making assertions—at least as we happen to find them. What is needed, is a third norm of *truth*, which he states as follows:

> Truth: If not-P, then it is incorrect to assert that P; if Not-P, there are prima facie grounds for censure of an assertion that P… [This] provides a norm of assertion which we take it that a speaker may fail to meet, even if she does not meet the norms of subjective assertibility and (personal) warranted assertibility. We are prepared to make the judgement that a speaker is *incorrect*, or *mistaken*, in this sense, simply on the basis that we are prepared to make a contrary assertion (170).

What would things be like without the third norm? Price thinks we can glean some picture by imagining a community of "merely opinionated asserters"—or MO'ans as Price calls them—whose linguistic lives were guided *solely* by the norms of sincerity and justification. The MO'ans would "criticize each other for insincerity and for lack of coherence, or (personal) warranted assertibility" yet they would not take disagreements to suggest that at least one speaker was mistaken (Price 2003, 172). Insofar as we can imagine such a community, it is hard to resist the temptation to say that their linguistic practices would be seriously *deficient*. Without the "friction" provided by the third norm, the MO'ans' discourse would simply be a "chatter of disengaged monologues." Perhaps even more seriously, Price argues, without the third norm, the idea of improving one's current commitments (or those of one's community) would be incoherent (174).[55]

Like Brandom's account of assertion, Price's functional account of truth satisfies the two criteria required by the global expressivist's 'upper tier.' It is both non-representational and it applies in a uniform way to the different "types" of assertion characterized by the local expressivist. That is to say, not only does Price's functional explanation of assertion and truth leave no room for substantial 'word-world' relations, but in so far as it explains such notions on the basis of their tendency to usefully coordinate beliefs, it remains agnostic, so to speak, about "what gets coordinated" (Price 2013, 50).

2.3.3 Reconstructing 'Representation'

I have mentioned that Price's vision is fundamentally *reconstructive*. Although there is a clear sense in which he is recommending that philosophers simply set aside certain semantic concepts *for various theoretic purposes*, he is neither suggesting that we should purge them from everyday usage nor that they are entirely devoid of theoretical interest. In particular, Price thinks that one advantage of global expressivism is that it offers us a perspective from which notions

[55] It is worth mentioning that some commentators have found Price's concern with the normative character of truth to put him in a somewhat awkward position regarding semantic deflationism. Deflationism, recall, is often taken to be the view that the equivalence scheme tells us everything that we could hope to know about truth. Price effectively agrees with the deflationist's rejection of the explanatory role of robust semantic properties and relations, but thinks that their usual *pragmatic* explications of the truth-predicate (i.e., that truth is a logical device for generalization) do not go far enough in making sense of the role that truth plays in human life. See especially (Price 1997; and Price 1998) for a discussion of this, as well as (Lynch 2015; Misak 2015).

like “representation” can be salvaged or reconstructed. This is one respect in which Price’s pragmatism more closely resembles Dewey or Peirce’s rather than, for instance, Rorty’s.[56]

For example, Price thinks that from the perspective of the global expressivist’s two “tiers” one can glean two very different functions of the concept of ‘representation’—which have typically been run together in a way that has caused great deal of philosophical confusion. For Price, the first sense, of “internal” or i-representation involves the kind of functional role that a linguistic item might play within some broader inferential system (Price 2013, 36). To take a simple analogy, just as a piece of chess represents, say, “the queen” by virtue of its use within a system of other pieces employed when playing a game, so too words and expressions “i-represent” by virtue of their function within a broader constellation of terms and expressions. The second sense, “e-representation” involves the idea of “environmental co-variance” (36). It is in this sense of ‘representation’ that one might speak when one says that a fuel gauge represents the level of fuel in a tank, or that a banana represents the presence of oxygen in the vicinity by turning brown. There are several advantages to be had by holding these notions apart. For one, it becomes apparent that they involve different notions of “external constraint.” Whereas i-representation involves “the kind of ‘in-game externality’ provided by the norms of the game of giving and asking for reasons,” e-representation involves a kind of environmental answerability (37-8). Moreover, whereas the notion of i-representation is meant to capture the idea at the heart of inferentialism, that semantic content depends on “the complex inferential relationships among and between linguistic items” (40), it is fundamentally a *separate* notion from the idea of

[56] Citing their personal correspondence, Price recollects that Rorty once commented: “My strategy is more slash-burn-uproot-sow-with-salt than yours” (Price 2013, 193).

"correspondence" which underwrites the notion of e-representation and its environmental constraint.[57]

One reason that such an appeal is important is that it allows the global expressivist to offer *some* concession to the traditional local expressivist's intuition that certain discursive practices are "genuinely representational" (Price 2013, 35). By reconstructing this distinction, Price can allow that *some* language games are more in the "e-representing" game than others (153). For instance, there may be cases in which particular scientific vocabularies rely more heavily on 'environmental tracking' paradigms than other vocabularies. This concession, however, comes with some important caveats. First, Price is clear that e-representation is not a semantic notion (37).[58] Second, because every (assertoric) discursive practice involves i-representations, in the same functional way, one cannot appeal to the notion of i-representation in order to draw distinctions between different vocabularies (38). Therefore, there is no way to reconstrue the bifurcation thesis in terms of representation.

2.3.4 Global Expressivism and the Problem of Completeness: Sociolinguistic Transformation Fallibilism, and Epistemically Virtuous Practices

In the remainder of this chapter I take up three objections which commentators have recently raised with respect to Price's global expressivism. These objections are all loosely related in suggesting that Price's position is incapable of adequately accounting for certain

[57] In availing himself to Brandom's inferentialism Price is committed to the usefulness of the notion of 'semantic content.' In earlier formulations of his position, he tended to lump this idea together with 'truth' and 'reference' as notions that pragmatists should reject when it comes to explaining meaning (See for instance, [Price 2004, 209]). Lionel Shapiro has taken issue with this tendency (Shapiro, 2014). In more recent publications, Price seems to have become more comfortable with semantic content.

[58] Price claims that "it is open to [the global expressivist] to take the view… that there isn't any useful external notion, *of a semantic kind*—in other words, no useful, general, notion of relations that words and sentences bear to the external world, that we might identify with truth and reference" (Price 2013, 37).

aspects of discursive practice. Ultimately, my aim in this section is to set the stage for a more sustained argument which I develop in the next chapter, which is that global expressivism has trouble accommodating an important kind of sociolinguistic transformation. As I go on to argue in Chapters Three and Four, the three criticisms discussed in this section can all be understood as evidence for an overlooked methodological limitation within subject naturalism which stems from the fact that its functional explanations of linguistic practices rely on generic, highly abstracted models. To anticipate, I shall argue that by turning to philosophical genealogy, subject naturalists can find the methodological elaboration required to overcome this limitation.

John Capps has taken aim at the notion of "i-representation," as part of a broader argument urging Price towards a "global pragmatism" which plays up the connection between meaning and action. Recall, that for Price, the notion of "i-representation" was governed by constraints *internal* to a particular language game. In contrast to representational accounts, according to which semantic constraints are "set by objects or what representations are supposed to represent" (Capps 2018, 74), Price's "i-representations" are governed by the norms involved in the game of giving and asking for reasons (Capps, 74). In this sense, semantic constraints are independent of individual speakers, but not somehow "external" to those linguistic communities of which they are a part. But, Capps argues, if this is correct, then it is unclear how one is to explain "systematic failures where a vocabulary or language game, despite being internally consistent, fails in its stated function" (75). The history of science provides an abundance of such cases (think, for instance of phlogiston theory), but examples abound in other areas of culture too.[59] Indeed, Capps terms this the "mumblety peg problem" in reference to a once popular

[59] Capps puts this problem as follows: "history is full of vocabularies that have fallen away not because they violate their own internal constraints, or because those internal constraints were internally incoherent or inconsistent, but because these vocabularies failed in comparison with other, incompatible, ways of doing what needed to get done" (Capps 2018, 76).

children's game in which players dropped pocketknives in close proximity to their own and others' feet. As far as games go, mumblety peg was "coherent" and presumably governed by a set of internal constraints. Yet, for this very reason, it seems that one would need to appeal to some set of external criteria by which to make sense of its ultimate demise: such as the fact that children seldom carry pocketknives these days, or that there are much better games to be played which are far less dangerous (81).

Capps suggests that Price can avoid the mumblety peg problem by broadening his notion of representation so as to introduce an element of external constraint into linguistic practice. His proposal, which is inspired by C. S. Peirce and Wilfrid Sellars, is to adopt a conception of "o-representation," according to which something counts as a representation "in virtue of its *operational* role in facilitating certain kinds of worldly interactions. O-representation gives priority to the action and conduct-guiding roles that concepts and statements play, based on the connection between one's cognitive architecture and the surrounding environment" (Capps 2018, 79). This friendly amendment would afford a perspective from which language games could be assessed not simply in terms of their internal coherence, but in terms of their success in guiding human conduct. This would allow Price to give "a fuller account of a language game's capabilities and shortcomings" (81). The reason that mumblety-peg and phlogiston theory both died out, on this picture, would be that "their o-representational content [was] thin, thinner than other theories" (81).

As we shall see, I ultimately agree with Capps that Price has trouble making sense of certain kinds of transformations of our linguistic practices, although for somewhat different reasons. While I think that he is right to suggest that an explanatory desideratum of global expressivism is that it should accommodate the action-guiding role of various uses of language, I doubt that Price needs the notion of "o-representation" to do this. Price could deny that the kinds of explanations Capps is looking for (that is, explanations of why some linguistic practices get replaced by others) require the availability of standards *external to the practitioners' perspectives*. As Capps himself suggests, part of the explanation of why mumblety peg (or, say, the effluvial theory of electricity and magnetism) was abandoned was because alternatives emerged (Capps 2018, 81). But this, of course, assumes that these alternatives could be formulated *as alternatives* within some vocabulary available to those who would come to adopt it. And while in some cases, appealing to the relatively superior action-guiding function of a discursive practice may suffice to explain why it was adopted, this hardly seems to be a necessary condition for such an explanation. Capps himself mentions that, at least in the scientific cases, practitioners often abandon one theory in terms of another through appeals to the well-known Kuhnian values of simplicity, explanatory power, fecundity, etc.. One wonders why appealing to these values could not just as easily provide the explanation that 'o-representation; is supposed to.[60] Given that such appeals may explain the kinds of changes in question, Capps would need to do more to show that they are ruled out by global expressivism.[61]

[60] Moreover, Capps doesn't acknowledge that different types of sociolinguistic practices might be subject to different kinds of internal (as well as external) constraints.

[61] A second way that Price might reply would be to insist that his broader, functional account of truth (rather than his narrower account of i-representation) can make sense of the fact that internally coherent practices sometimes get abandoned. For instance, a crucial premise of his evolutionary explanation of why human beings employ the truth predicate is that doing so helps improve beliefs. Given sufficiently complex linguistic practices, as well as norms encouraging disagreements to be resolved through argument, it isn't clear that Price requires a further set of external constraints to explain how proposals for novel language games might emerge, give rise to dispute, and occasionally result in the adoption of new practices. If, however, and as I ultimately suggest in Chapter Three, Price's characterization of these norms (of disagreement) somehow *precluded* new ways of speaking from being integrated into existent practice, this would be a problem for him as it would seem to render linguistic transformation unintelligible.

In another vein of critical engagement, Henrik Rydenfelt has recently argued that Price overlooks the possibility of a non-representationalist scientific realism, rooted in the insight that only *some* of our discursive practices involve commitment to a kind of fallibilism. Here, part of Rydenfelt's complaint is that Price's account of assertion is too coarse-grained.[62]

Following Peirce, Rydenfelt argues that one hallmark of science, which distinguishes it from other ways of settling opinions, such as appeals to authority, is that the former is characteristically aimed at "ascertain[ing] how things truly are independently of our opinions" (Rydenfelt 2019). This "fundamental hypothesis" entails a number of ontological commitments. For instance, that engaging in scientific practice requires "that there is an independent reality" and that "real things… affect us causally through perception, causing us to form judgments" (Rydenfelt 2019). This is because, as Rydenfelt explains, the "reasons given for or against a belief often make reference to reality" (Rydenfelt 2019). Crucially, however, a corollary of the fundamental hypothesis is *fallibilism,* understood as the "view that any of our opinions may be mistaken."[63] This, in turn, applies to the ontological commitments themselves—hence Rydenfelt's label of *hypothetical* realism.

Rydenfelt argues that the availability of hypothetical realism constitutes a problem for Price. This is because the latter's theory of assertion is ill-equipped to account for the fact that fallibilism seems to be constitutive of some discursive practices, but not of others. Although Price provides enough resources to ensure the publicity of truth—in so far as disagreements tend

[62] Although this is a very different kind of criticism than the one presented by Capps, they are similar in so far as they target the 'global' side of Price's project, namely his account of assertion.

[63] More precisely, fallibilism is understood as a modal thesis according to which for any proposition, p, "either it is possible that (we believe that p but it is the case that not-p), or else, it is possible that (p but we do not believe that p)" [Rydenfelt 2019]).

to ensure that one speaker is wrong—this does nothing to guarantee that "truth is independent of the opinions of anyone" (Rydenfelt 2019). After all, Rydenfelt contends, "a group of religious fundamentalists, say, may subscribe to Price's third norm, thereby criticizing each other for speaking what is not true by their own lights. Nevertheless, they still maintain that the Holy Book is infallible, at least concerning some issues" (Rydenfelt 2019). In some cases, such a practice of "settling opinions" may strike us as objectionable; although in others it might seem perfectly reasonable. In any case, clearly, we would *not* want to say that it is a *fallibilistic* practice. Rydenfelt's point, however, is that Price leaves us with no way of drawing the line between those practices which characteristically involve fallibilism and those that do not.[64]

Like Capps, Rydenfelt ultimately thinks that global expressivism can be rescued by amending Price's characterization of representation. Price sometimes intimates that the notion of e-representation could provide a means of demarcating scientific vocabularies from non-scientific ones. Skeptical about the viability of this approach, Rydenfelt suggests that the global expressivist instead "reinterpret e-representation in terms of the aim of scientific project" (Rydenfelt 2019). For Peirceans like Rydenfelt, this means that we distinguish scientific practices from non-scientific ones on the basis of the idea that it is only for the former that opinions are to be settled independently of what we happen to think. Scientific practices would, then, involve a fallibilistic commitment to an independent reality—which amounts to a kind of

[64] Rydenfelt goes on to explain, "Why this distinction does not emerge in Price's discussion seems reasonably clear: fallibilism is difficult to detect in a consideration of assertoric practices in general. Rather it surfaces in our practices of settling and justifying opinions. Price's norm, as advertised, is a norm of assertion; but how disagreements are to be resolved—how our opinions are to be settled and justified—is an issue on which that norm is not intended to bear" (Rydenfelt 2019).

realism. And although Price has typically shunned this kind of characterization of his project, Rydenfelt insists that he has every reason to accept it.[65]

In some important respects, I agree with Rydenfelt's critique. That is, I do not think that Price's account (at least as it stands) has the requisite resources to draw the kind of distinction between fallibilist and non-fallibilist practices on which Rydenfelt insists. However, I am not sure that Price would agree with the route he takes to arrive at this conclusion. First, I imagine that Price would be suspicious of the notion of "reality" that the hypothetical realism is supposed to rely on. This is because I am not sure that Price (or anyone for that matter) should concede Rydenfelt's claim that our "reasons given for or against a belief often make reference to reality."[66] Even if, in practice, people did tend to make such an appeal, it is not clear that *conceptually* it is robust enough to serve as a useful justification.[67] Second, Rydenfelt seems to

[65] As Rydenfelt explains, "The divide between what falls under the scope of science and what does not is contingent: it depends on our varying practice of settling and justifying opinion. The epistemic thesis [that only scientific theories are at least approximately true of reality] is thus replaced by a far more modest understanding of the aim of the scientific practice. Nevertheless, that practice entails an ontological thesis: hypothetical realism, the assumption of an independent reality. Accordingly, if refashioning E-representation along these lines is the most feasible option for Price, the Carnapian rejection of "both the thesis of the reality of the external world and the thesis of its irreality" is not quite successful: Price is a realist, after all" (Rydenfelt 2019).

[66] Suppose there is some matter about which Rydenfelt and I want to settle things independently of either of our opinions, say, whether a given solution is acidic or basic. We might perform a litmus test in order to find out. Should we disagree about what to conclude from the test, and therefore feel the need to justify out beliefs to one another, there are many reasons that we might offer to one another—for instance, about the results of the test, our broader understanding of chemic theory, etc. But at no point would it seem that an appeal to something called 'reality' would either be *required* or *convincing.*

[67] Diana Heney is critical of Price's account of assertion for much the same reason that Rydenfelt is. Namely, that it does not leave room for the notion of 'reality' as a regulative assumption within our epistemic practices. On the Peircean model that Heney recommends, "The method of inquiry relies on the supposition that there are real things with which we want our beliefs to be in harmony, *but Peirce does not claim anything specific for the content or nature of that reality*. It is the idea of reality in the role of a working hypothesis upon which Peirce relies, and upon which he thinks we must rely if we are to make sense of our practices of inquiry and assertion" (Heney 2015, 507, italics added). But if 'reality' is supposed to be a hypothesis about which one is not claiming any specific content, or to which one is not attributing any nature, it is hard to see what difference such a hypothesis would make in our epistemic practices.

be saying that fallibilism is a defining characteristic of science. But this seems to be both too broad and too narrow.[68]

Finally, allow me to consider a third criticism which also targets the explanatory adequacy of Price's project. Cesare Cozzo argues that Price's account of truth fails to deliver a faithful picture of epistemically virtuous practice (Cozzo 2012). To make this point, he proceeds through a series of thought experiments envisioning different linguistic communities whose practices resemble ours in certain respects but which differ in other crucial ways. The aim is to throw light on the role that the truth norm plays in our own lives by offering us a glimpse of how things would be without it. The first imagined community (hailing from what Cozzo calls 'Laconia') find themselves in the same situation as Price's "merely opinionated asserters" or MO'ans. That is, speakers in Laconia are governed by norms of sincerity and justification, but they remain "indifferent to the fact that their own announcements and those of their fellow-speakers often disagree" (Cozzo 2012, 60).[69] Like Price, Cozzo denies that these language-users could be said to possess *our* notion of truth. But unlike Price, he denies that simply adding a further normative component to the Laconians' discursive practice—namely one according to which conflicts between sincere, justified claims would produce disagreements leading to some sort of resolution—can yield a satisfying model for the normative character of truth. He makes this point by envisioning a second community, the 'Erisians,' who differ from the Laconians only to the extent that they "disapprove of those with whom they disagree and do all they can to

[68] Rather than attempt to work up an account of fallibilism out of an ontological thesis as Rydenfelt does, I suspect that Price would be more inclined to regard it as an intellectual virtue that appears in some practices but not in others. And then to explain (and in some cases justify) the existence of this virtue by reference to the benefits it confers to individuals and communities who adopt it. In any case, this is roughly what I shall propose in the following chapter.

[69] They also possess the terms 'true' and 'false,' which they employ in the same generalizing and disquotational ways that we do (Cozzo 2012, 60).

resolve the disagreement" (61). Whereas for the Laconians, disagreements are inert, the Erisians are sensitive to and motivated by the fact that one's social status is often a function of whether others accept one's claims. The Erisians view disagreements as struggles for power; and, hence, ultimately aimed at persuasion. Given their view that "in debates the only aim is to win, by all means, at any cost" (61), members of this community (at least the most socially successful ones) are well-equipped with the kinds of tools that Plato's sophists specialized in wielding—the ability to persuade through subtle attacks on an opponent's character, equivocation, and knowing how to advantageously suppress evidence. In short, the Erisians engage in what Cozzo calls an "agonistic-persuasive practice" (61).

The crucial question is whether such a practice is sufficient for having the norm of truth. Price's account seems to carry the unwanted implication that it is, and for Cozzo, this is demonstrative of its inadequacy. While the Erisians care about winning arguments, nothing about their discursive practices suggests that they care about truth. What this suggests is that there are different ways of resolving disagreements, some more virtuous than others. The problem with Price's account of assertion is that it seems incapable of making sense of this distinction. In order to possess a truth norm that we could recognize as our own, participants in a linguistic practice need to do more than just treat their disagreements as problems in need of resolution. At least some of the time, they need to treat their disagreements as problems that ought to be resolved "by epistemically virtuous practices" (Cozzo 2012, 65). What do such practices look like? For Cozzo, an epistemically virtuous practice demands that its practitioners possess certain behavioral traits including (but not limited to): a willingness to hear both sides, open-mindedness, impartiality, care and sensitivity to detail, attention to evidence, willingness to question assumptions and to offer and ask for reasons, intellectual curiosity and courage, as well as readiness to revise one's assertions (62).[70] While Price may have identified an important norm associated with assertions, he does not appear to have captured *our* norm of truth.

Like Capps, Rydenfelt, and Cozzo, I find Price's account of assertion to be well motivated, but ultimately unsuccessful in its current form. Like Capps, I think that the account fails to make sense of sociolinguistic change (though for different reasons). Moreover, like Rydenfelt and Cozzo, I believe that Price's account of assertion needs to be broadened to include a wider range of norms *in addition to* sincerity, justification and truth. I develop these criticisms in greater detail in the next chapter.

2.4 Conclusion

In this chapter I have offered a detailed overview of Huw Price's global expressivism. Price's attempt to reconstruct philosophical naturalism along *subject naturalist* lines represents a radical way of thinking through some of the most pressing issues in contemporary metaphysics. Moreover, the success of Price's project hinges on its ability to overcome two sets of objections. The first, corresponding to what I called the problem of transcendence has to do with the question of whether subject naturalism can make sense of normativity in a way that does not threaten to undermine our commitments to certain values. The second, corresponding to the problem of completeness, has to do with whether the account of truth and assertion which serves as the "upper-tier" of Price's global expressivism can make sense of important elements of linguistic practices, such as the importance of fallibilism in certain contexts, the epistemic virtues, as well as the possibility of sociolinguistic change. Having drawn attention to these

[70] In generating this description, Cozzo is drawing from Michael Patrick Lynch's discussion of the value of truth (Lynch 2004).

potential problems and their significance given Price's overall project, I now turn to the task of showing how subject naturalism can be amended so as to avoid them.

CHAPTER 3

SUBJECT NATURALISM AND THE PROBLEM OF SOCIOLINGUISTIC REVOLUTION

3.1 Overview and Aims

When successful, subject naturalism can serve as a powerful tool for demystification. This is not to say that it can reveal a hidden reality that lies beneath the depths of appearance; but that it can, to borrow a key Wittgensteinian phrase, offer us a perspicuous or surveyable depiction of our practices.[1] For philosophers held captive by the philosophical picture that is object naturalism, this is often just what is needed. After all, what could be more *human* than moral or aesthetic value, mindedness, or meaning? What drives the contemporary metaphysician is the unsettling thought that our concepts have somehow taken on a strange aura insofar as they cannot—like misshapen puzzle pieces—find their place within the natural world. By offering explanations of the functional role that our concepts play within human life, and by drawing our attention to the needs and exigencies to which our language games respond, subject naturalism can be a potent form of philosophical therapy. But it is important to see that it can offer more. Even those who are not gripped by the object naturalist's metaphysical pretensions may come to lose sight of the point of their practices. One of the claims developed in this chapter and the next is that subject naturalistic inquiry can lead us to see our practices in new and unexpected ways,

[1] In *Philosophical Investigations* Wittgenstein writes that, "A main source of our failure to understand is that we don't have *an overview* of the use of our words. — Our grammar is deficient in surveyability. A surveyable representation [übersichliche Darstellung] produces precisely the kind of understanding which consists in 'seeing connections'" (Wittgenstein 2009/1953, 54).

and that in doing so it can change our attitudes towards those practices. But understanding how they do this is by no means straightforward. It remains an important philosophical task worthy of attention.

In the previous chapter, I suggested that Huw Price's global expressivism faces two significant clusters of problems. First, as several commentators have pointed out, there are shortcomings in his account of assertion and truth. In particular, it is unable to make sense of important aspects of discursive practice, such as sociolinguistic revolution (Capps 2018), fallibilism (Rydenfelt 2019), and the distinction between epistemically virtuous practices and agonistic-persuasive practices (Cozzo 2012). The second set of challenges involved subject naturalism's relationship to normativity. While Price is clearly opposed to any form of scientism—according to which the sciences are afforded a foundational place within culture—his tendency to identify "the natural" with "the scientific" threatens to render his subject naturalism too narrow. On the one hand, some commentators have argued that Price is unable to make sense of irreducible normative notions such as meanings, values, or reasons, whose first-personal nature prevents them from being studied exhaustively by scientific investigation (Macarthur 2014). On the other hand, we saw that Price's conception of naturalism is threatened by a potentially more serious worry, which is that it could end up undermining the very normative commitments that it purports to understand or explain (Redding 2010). As we shall see, in some cases, the availability of the detached, reflective perspective presupposed by subject naturalism can have a destabilizing effect on one's commitments. I call this the stabilization problem.

Both sets of challenges—if successful—would suggest that Price's employment of subject naturalism has failed in its demystifying ambitions. While it may have dispelled certain unproductive metaphysical urges, it has only done so at the expense of leaving us with new perplexities. How can we make sense of linguistic revolution, the role of fallibilism or the epistemic virtues within our discursive practices? Are we able to sustain our normative commitments to those features of our linguistic lives targeted by subject naturalistic inquiry? The aim of this chapter and the one following it, is to explore these criticisms in greater detail and to suggest and defend some possible avenues that a pragmatist—especially one sympathetic to naturalism and metaphysical quietism—can take to overcome them.

In this chapter, I draw attention to an important deficiency in Price's account of assertion by way of considering of the phenomenon of sociolinguistic revolution. Any adequate account of assertion must be able to accommodate the fact that, linguistic communities occasionally adopt radically new ways of thinking and talking about themselves and the world. Paradigmatically, these forms of sociolinguistic transformation occur in moral and scientific revolutions. My claim is that if one accepts Price's view of the normative character of truth, it is impossible to make sense of how radically new contributions to moral and scientific knowledge come to be integrated into standard discourse. Put in Kuhnian terms, the problem is how—given Price's account of assertion—to make sense of the possibility of shifting from "abnormal" science or discourse, to "normal" science or discourse.

Ultimately, I believe that Price can accommodate the phenomenon of sociolinguistic revolution by expanding the methodological resources of subject naturalism. In particular, I shall argue that the problem of sociolinguistic revolution requires the subject naturalist to broaden their account of assertion to include a wider constellation of values in addition to sincerity, justification, and truth (which are to be understood as historically variable and contingent). In the next chapter, I will return to take up this explanatory deficiency in Price's account. By drawing from recent work in philosophical genealogy, especially that of Bernard Williams and Matthieu

Queloz, I will argue that the subject naturalist requires greater methodological attention to the de-idealizing, historical components of their genealogical explanations.

As we saw in the previous chapter, Price often conceives of the positive project of his subject naturalist stance—global expressivism—as making use of what he refers to, but with precious little explanation, as *genealogy* or *linguistic anthropology*.[2] The overarching conclusion of this chapter and the next is that Price, and more importantly those who wish to take up and expand his valuable insights would do well to reflect further on the methodological demands and possibilities of genealogy as a kind of philosophical explanation. Although I focus primarily on the work of Bernard Williams, I do not intend to rule out any of the rich and varied intellectual resources of other philosophical contributions to genealogy, such as those of Friedrich Nietzsche or Michel Foucault. Or, as I suggest at the end of this chapter, one might even look to classical pragmatism—in particular, to the work of William James and John Dewey—as providing a model of a kind of philosophical genealogy from which subject naturalists might draw.

3.2 The Problem of Sociolinguistic Revolution

In the previous chapter, I claimed that if Price's account of assertion were correct, it would be difficult to make sense of radical linguistic revolution. That is, one would have trouble explaining how novel ways of speaking or new conceptual developments become integrated into the standard discourse of a linguistic community. I shall now attempt to flesh this criticism out in

[2] See Macarthur and Price (2007, 231); Price (2010, 320; 2013, 58, 59, 61, 62). I have not been able to find any passages in Price's work in which he attempts to differentiate these two terms. In one passage, Price describes an "expressivist genealogy for causation and other modal notions" as involving "a scientific account of a particular aspect of human linguistic and cognitive practice, explaining its origins in terms of certain characteristics of ourselves, as epistemically limited creatures, embedded in time in a particular way" (Price 2013, 61). This brief characterization leaves many questions unanswered: is genealogy supposed to be a form of evolutionary biology? Does it require attention to human history? What are the criteria by which to distinguish between successful and unsuccessful genealogies? In this chapter I shall suggest that Price should look to recent work on philosophical genealogy for answers to these questions.

more detail by drawing on some helpful terminology proposed by T. S. Kuhn and Richard Rorty. Ultimately, I shall propose that Price's account of assertion needs to be broadened to include a wider set of norms and values in addition to sincerity, justification, and truth.

In *The Structure of Scientific Revolutions* Kuhn employed the notion of "normal science" to capture the idea that, aside from those exceptionally rare "revolutionary" or "abnormal" cases in which scientists must choose between competing theories, the vast majority of scientific research proceeds in light of a background of shared criteria for determining which problems mattered and how they ought to be resolved.[3] This approach inveighed against the established approaches to the philosophy of science in at least two respects. First, it set aside attempts to construct an ideal model or rational reconstruction of scientific reasoning and instead tried to understand scientific inquiry as it is actually practiced, both historically and in contemporary contexts. Second, it undercut the image of scientists as the paragons of critical thinking or intellectual courage who, as Karl Popper thought, went around boldly aiming to falsify their theory whenever such an opportunity presented itself. [4] Instead, Kuhn represented scientists in a more prosaic light as "puzzle solvers" who were typically indoctrinated into the methods and values of their practice.

In the final part of *Philosophy and the Mirror of Nature,* Richard Rorty suggested that we can think of *any* area of culture involving cooperative human activity along similar lines. *Normal*

[3] Kuhn writes that normal science means "research firmly based upon one or more past scientific achievements…that some particular scientific community acknowledges for a time as supplying the foundation for its further practice" (Kuhn 1962, 10). By contrast, scientific revolutions occur when existing theories encounter enough anomalies that practitioner begin to search for new "a new set of commitments, a new basis for the practice of science" (6). Revolutions are "the tradition-shattering complements to the tradition-bound activity of normal science" (6).

[4] Popper expresses this conception of scientists in his discussion of falsifiability in *Conjectures and Refutations: the growth of scientific knowledge* (Popper 1963, 33-39).

discourse (like normal science), he urged, occurs whenever practitioners agree on the types of intellectual contributions that count as relevant, and whenever there is consensus regarding the criteria by which such contributions are to be judged. *Abnormal discourse*, by contrast, occurs when such consensus is lacking (Rorty 1979, 320). Those who engage in normal discourse share a common vocabulary and wield concepts whose meanings are not typically subject to radical interpretive disagreements. This allows for a key feature of normal discourse, which is that it contains a relatively stable set of criteria for assessing knowledge claims. By contrast, the knowledge claims constitutive of abnormal discourse will lack a set of stable inferential connections, and will employ concepts whose meanings are relatively unstable. From the perspective of normal discourse, abnormal knowledge claims will appear patently false.

Conceiving of Kuhn's distinction more broadly allowed Rorty to insist on a strong anti-foundationalism, while avoiding the charge of irrationalism. Notions of objectivity and rationality need not be thrown overboard. They still find a place within any sociolinguistic practices within which shared criteria are in place; but they are not grounded in anything timeless or eternal that stands beyond those contingent practices. This move was part of Rorty's attempt to deflate the pretensions of epistemically-centered philosophy, construed as the search for a "permanent neutral framework" within which to adjudicate between any competing intellectual contribution. On Rorty's "post-epistemological" vision of philosophy, there is no vantage point from which to insist on a kind of cultural hierarchy in which some areas are to be regarded as "more rational" or "more in touch with reality" than any other—unless, of course, this is taken to be a claim about the relative presence of shared norms or agreed-upon procedures for resolving disagreements.[5] Nor are there any transhistorical criteria on the basis of which large-scale cultural revolutions could be assessed (Rorty 1979, 332).[6] None of this entails that judgments about intellectual progress are impossible, only that such judgments will always be retrospective, an attempt to say how things stand in light of one's current standards (1979, 332).

Following Rorty, abnormal discourse (like revolutionary science) is often an indispensable growing point of language. The process by which radically new ways of speaking are taken up in standard usage is one by which they begin to acquire a new set of inferential connections. For these new contributions to our vocabularies become adopted, is just for them to occupy a more stable place within normal discourse, thereby (potentially) standing in justificatory relations with other statements (Rorty 1991a, 171). In contrast to "perception and inference" which allow us to acquire new beliefs by altering the truth values of previously accepted sentences, sociolinguistic revolutions occur when logical space itself undergoes a radical change, by "expanding our repertoire of sentences" (Rorty 1991b, 12).

On this picture, one can distinguish between two distinct kinds of sociolinguistic change. For purposes of clarity, I shall refer to the first as 'improvement' and the second as 'revolution.' Drawing new inferences and having novel perceptions can lead us to improve our beliefs about the world. But those processes cannot, by themselves, bring about an alteration of the logical space of reasons. That is, they cannot revolutionize normal discourse. For thinkers like Rorty, this is an important reason for eschewing teleology in history. Radical sociolinguistic change can

[5] Here, Rorty's position aligns with Price's functional pluralism about linguistic frameworks, in which science is just understood as one discourse among many.

[6] From the perspective of Rorty's epistemological behaviorism, "nothing counts as a justification unless by reference to what we already accept, and … there is no way to get outside our beliefs and our language so as to find some test other than coherence" (Rorty 1979, 178).

be understood as facilitating moral and scientific progress, not by allowing us to more accurately represent the world or express our true natures, but by "changing the way we talk, and thereby changing what we want to do and what we think we are" (20).

Although Price's account of assertion can help explain what I have called "the improvement of beliefs," whereby one draws new inferences or gains new insight through perception, it seems to me that it threatens to render unintelligible linguistic revolution, whereby the space of reasons becomes radically altered. We can begin to see why by considering that contributions to abnormal discourse will almost always *patent falsehoods* from the standpoint of the semantics of standard usage. The extent to which expressions are true from our present standpoint depends on their having become successfully integrated into standard usage. The question I want to raise is whether Price's view even allows for the possibility of such revolutions to occur. My claim is that it cannot.

To see why, recall Price's third norm of assertion—the norm of truth (connected to the other two norms of sincerity and justification)—which I cited in the previous chapter:

> Truth: If not-P, then it is incorrect to assert that P; if Not-P, there are prima facie grounds for censure of an assertion that P (Price 2003, 170).

In defending this norm as an indispensable component of assertion, Price takes it to have two important features. First, he contends that truth is a *default* norm of assertion in the sense that it is always engaged until speakers can be brought to accept that disagreement is unproductive. Second, the norm is supposed to be *ubiquitous* in the sense that it tends to occur in all assertoric practices—it is precisely this feature of the norm of truth that allows the global expressivist to avoid the bifurcation thesis by jettisoning the notion that certain domains involve 'genuine' assertions whereas others do not.

If Price is correct about the default engagement and ubiquity of the third norm, then *prima facie,* the appropriate reaction to an abnormal utterance P, would be for an audience to regard its utterer as *incorrect* and to try to get them to assert that not-P. After all, the normative function of truth in our linguistic practices is precisely to encourage disagreement between speakers—this function is, as I discussed in the previous chapter, thought to be advantageous because it ultimately leads members of a linguistic community to pool their cognitive resources (Price 1988, 145). But given that, from the perspective of standard usage, abnormal utterances are *false*, it follows that when confronted with such utterances, competent speakers would characteristically be inclined to voice their disagreement through argument. Price's norm entails, therefore, that our default reaction to hearing an abnormal claim would be to engage in a reasoned argument with the person who uttered it, with the aim of getting them to change their mind.

The question I am considering is whether this picture of assertion can allow for the process through which abnormal discourse becomes normalized, thereby transforming normal discourse. It might seem like it can. After all, couldn't Price claim that it is through *arguments* these radical knowledge claims are brought to earn their keep within the confines of standard use? If Kuhn's general idea is correct, this cannot be the case. This is because abnormal discourse—by definition—falls outside the bounds of predictable linguistic behavior and, therefore, lacks a set of stable inferential connections. This makes abnormal assertions unsuitable candidates as premises and conclusions in inferences. But if they cannot play such a role in arguments, this would seem to preclude the possibility that their integration into normal discourse—that is to say, their acceptance as being *true*—could be gained through argument alone. Therefore, given the default engagement of Price' third norm, it is hard to see how a scientific or moral 'revolution'—understood as the enlargement, or alteration of logical space—could arise.

At this point, Price could reply that abnormal uses of language are precisely those kinds of cases in which the third norm is likely to disengage, where disputes are liable to "evaporate."[7] Our vast repository of discursive experience has taught us that some disagreements are, in general, not worth having. Even the Scholastics, for whom few topics seem to have escaped philosophical discussion, were privy to the insight that *de gustibus non est disputandum*.[8] At some point along the way, Price might contend, our linguistic forebearers recognized that abnormal assertions were not the kinds of utterances worth arguing about.

There are at least two problems with this response. The first is that it seems *ad hoc* relative to Price's more general strategy for making sense of evaporative disputes, which premises a linguistic community's recognition that certain classes of assertions produce no-fault disagreements on some relevant features of those speakers or their environments (or what in later works, Price refers to as their practical situation).[9] To take one of Price's own examples,

[7] In *Facts and the Function of Truth* Price draws an important distinction between *evaporative disagreements* and *substantial disagreements*. The latter occurs "so long as the parties concerned regard each other as mistaken, and [disagreement] evaporates when such evaluations are no longer felt to be appropriate" (Price 1988, 161). He goes on to consider several ways in which disagreements may come to evaporate. On the one hand, there can be features of "the context or the conversational role of the judgments concerned" (195). Probabilistic, and conditional claims are good examples of this because of their sensitivity to the contextual nature of evidence (195). On the other hand, disputes can evaporate because of idiosyncrasies involving the speaker. An example of this would be subjective judgments about secondary qualities, or when a speaker grasps the meaning of a term in an idiosyncratic way (195-97). What is important for my argument is that Price takes all of these sources of evaporative disagreement to be explainable from the perspective of linguistic function. That is, whenever we encounter an area of discourse prone to evaporative disagreements, we can, in principle, explain this feature by appealing to some supposed function that it is supposed to serve in human life.

[8] Tastes are not to be disputed.

[9] Traditionally, philosophers have tried to explain why certain areas of discourse give rise to evaporative disagreements by appealing to the distinction between *factual* and *non-factual discourse* (i.e., disputes about tastes tend to evaporate *because* there are no facts of the matter about tastes). But Price thinks "that this gets things back to front", suggesting instead "that judgements of a certain kind are factual just to the extent that we do treat all disputes involving such judgements as substantial. Evaporative disagreement marks the limits of what we might characterize

speakers often make conflicting probabilistic assertions about the likelihood for some event. But, for the most part, such disagreements are afforded a certain degree of tolerance whenever speakers recognize that their justification is relative to subjective sets of evidence (Price 1988, 159-161). Suppose for instance, that Smith thinks that there is a 50% chance that a coin-toss will come out 'tails.' If Jones were to know that the coin was in fact biased and more likely to come out 'heads,' he would presumably still regard Smith's judgment as a reasonable one. Price's point is that as the kinds of creatures who plan and act within a world of uncertainty, we have come to recognize that our probabilistic judgements are often perspectival. It is our experience of this feature of our practical situation (along with our need to coordinate our behavior with others) that explains our preparedness to disengage the norm of truth when we encounter speakers whose probabilistic assertions differ from our own (160).[10] Is there then, an analogous feature of our practical situation that explains why abnormal assertions, considered as a distinct class of utterances, are liable to produce a disengagement of the third norm? It does not appear that there is. One reason for this is that it seems hard to identify some generic human need or problem to which abnormal speech—understood as a univocal linguistic phenomenon—can be reasonably understood as a response. That is, even if it did make sense to categorize abnormal judgments in

as the factual patter of usage" (161). The philosophical task, on this picture, is to explain why speakers tend to treat certain disputes as evaporative *given general features about them and their environments.* What I am claiming is that there is no such "general feature" to which we can appeal to explain why abnormal utterances (as opposed to, say, probabilistic judgements) tend to give rise to evaporative disagreement.

[10] It is important that these same considerations do not apply to every kind of assertoric utterance. Price makes this point by contrasting the kind of situation just describe with one in which we might be inclined to use the words 'true' and 'false." He writes, "In fading light, a guest mistakes my rabbit for a rat. 'There's a huge rat in your garden!', he cries. I reassure him, and politely excuse his mistake: 'That rabbit *does* look ratty at this time of night.' Excusing the mistake is not the same as acknowledging that no mistake has been made, however" (163). By contrast, "The special character of probabilistic judgements shows up particularly in the following. It can be reckoned correct to make a probabilistic judgement even if one knows of the existence of evidence which, if one had it to hand, might make one revise that judgement. For example, a doctor might say, "You are probably not infections, but if we had your test results we would have a better idea" (163). However, the same does not hold for non-probabilistic judgements. For instance, it would not make sense for a doctor to say 'You are not infectious, but if we had the test results we would have a better idea' (163).

the same unified way that we do with, say, probabilistic judgements, there does not seem to be any generic aspect of our practical situation that would explain why we employ the former kinds of claims, in general.[11] By Price's own lights, no-fault disagreements arise as the result of language users having taken a particular practical stance *as the result of exigencies of their practical situations* (e.g., as the possessors of tastes which are perspectival, as creatures who must decide under conditions of uncertainty), but abnormal utterances do not seem to admit of a similar kind of explanation.

And even if such an explanation were forthcoming, it is still doubtful whether Price's account could explain how an audience could come to accept or adopt radically new ways of speaking. A second problem stems from his characterization of such no-fault, or evaporative disagreements as forms of conversational *disengagement*. That is, as instances in which speaker and audience no longer take their disagreement to "matter." The problem is that in order for unfamiliar uses of language to work their way into standard usage, it seems plausible that an audience would need to take up a certain set of attitudes towards those utterances—perhaps a kind of openness or curiosity. There needs to be other norms in addition to sincerity, justification, and truth within the discursive practice to which one can appeal in order to make sense of how an audience might somehow take seriously or embrace new uses of language.

I can illustrate this point by considering two paradigmatic agents of moral and intellectual change who embody the account of sociolinguistic revolution I have been describing. I shall refer to these agents as moral prophets and revolutionary scientists. Moral prophets are social

[11] As we have just seen, probabilistic judgments can be plausibly understood as arising from the fact that human beings make decisions under uncertainty. But there are many different reasons people will have for trying out new ways of speaking. Perhaps it might be thought that abnormal discourse arise out of some basic human need for novelty.

critics who seek to transform the institutions, practices, attitudes, or sentiments of their communities, often in the face skepticism or even hostility. As Mark Johnson has put it, moral prophets "demand conscientiousness about moral matters we do not wish to behold, and they challenge us to change our ways" (Johnson 2014, 106). Revolutionary scientists are individuals or groups whose research takes place beyond the stable ground of Kuhnian normal science and its agreed-upon methods and cognitive criteria. They envision new possibilities for future investigation and lay the groundwork for the normal science of a future generation.

A feature common to both moral prophets and revolutionary scientists is their frequent reliance on abnormal discourse. While by no means a necessary condition for social revolution, it seems hard to deny that in certain cases, moral and scientific revolutions are dependent on widespread changes in a linguistic community's shared vocabulary. Moral prophets, for example, may coin altogether novel expressions in order to call attention to hitherto overlooked forms of discrimination or oppression. Or they may suggest radical extensions of familiar moral terms—such as 'suffering,' 'humiliation,' 'person,' or 'care'—to cases where such terms previously were deemed inapplicable. Similarly, on the Kuhnian view which I described above, a scientific revolution "involves the replacement of an accepted scientific lexicon or taxonomy with a new one" (Wray 2011, 15). While revolutionary scientists may often employ the same terms as their pre-revolutionary or "normal" counterparts, the meaning of such terms will often differ radically owing to a fundamentally different usage within the new framework.

To the extent that moral prophets and revolutionary scientists do rely on abnormal discourse to facilitate sociolinguistic revolution, both agents would seem to face a similar kind of situation, which to modify a term coined by Susan Dieleman, can be characterized as one of *semantic exclusion*.[12] From the standpoint of standard linguistic usage or received scientific practice, prophets and revolutionaries will often be perceived as nonsensical, irrational, or crazy, insofar as their discursive contributions do not conform to received patterns or paradigms. This is reflected in the fact that most changes in moral sentiments and attitudes which could be considered instances of moral progress—at least from the perspective of inhabitants of liberal democracies—can be described in terms which, though seeming perfectly reasonable from one's present perspective, would have surely appeared incomprehensible to previous generations. "Once, for example," writes Richard Rorty, "it would have sounded crazy to describe homosexual sodomy as a touching expression of devotion" (Rorty 1998, 204), and that such a description is "now acquiring popularity" should remind us that today's vocabulary of moral common sense may very well have been a nonsensical-sounding assertion in the mouth of yesterday's moral prophet.

The fact that agents of sociolinguistic revolution often face situations of semantic exclusion not only underscores that their radical knowledge claims are unlikely to be adopted through rational argument alone, but it also suggests that the success and perhaps even the very existence of prophets and revolutionaries depends on practices involving norms and values in addition to those of sincerity, justification, and truth. Because their calls for revolution are often couched in terms which fall outside of normal discourse, even the sincerest agent of sociolinguistic revolution is unlikely to succeed if the only means of securing uptake is through justification. Insofar as prophets and revolutionary intellectuals do succeed in affecting change, it hardly seems to be on account of their sincerity or their ability to justify themselves. Rather, a

[12] Susan Dieleman has argued that Rorty's account of sociolinguistic revolution is especially apt in making sense of the problem of *epistemic exclusion*. This occurs in situations in which those who are "unjustly excluded from invoking" dominant epistemic norms and practices find themselves incapable for challenging those norms, precisely because of the fact that they are excluded from invoking them" (Dieleman 2012, 90).

large part of their success depends on their having been taken seriously, which is to say that they have been engaged with sympathetically and with an open imagination. In other words, the likelihood that a moral or scientific community will be amenable to sociolinguistic revolution through prophetic moral agents or revolutionary scientists is a function of the availability of widespread values and dispositions within that community. These resources include, but are not limited to open-mindedness, care and sensitivity, willingness to question assumptions, imagination, intellectual curiosity and courage, as well as readiness to revise one's commitments. In short, agents of sociolinguistic revolution are only likely to succeed to the extent that they are embedded in the kinds of practices resembling those which Cozzo and Rydenfelt characterize as epistemically virtuous or fallabilist, which we saw in Chapter Two.

With this illustration in mind, I can now restate my objection to Price as follows: by limiting his account of assertion exclusively to the norms of sincerity, justification, and truth, he lacks the explanatory resources to make sense of sociolinguistic revolution. I arrived at this conclusion by way of a dilemma. On the one hand, if one accepts Price's characterization of the truth norm as being both *ubiquitous* and *engaged by default*, then it is mysterious how instances of abnormal discourse could ever be accepted. This, I claimed, was because such contributions are almost always patent falsehoods, which implies that on Price's model they would automatically give rise to reasoned disagreements. But since abnormal assertions lack a stable set of inferential connections from the standpoint of normal discourse, their acceptance cannot be gained through argument alone. On the other hand, I argued that even in those cases in which Price's truth norm might be expected to disengage—producing what he calls "evaporative disagreements"—it does not seem that the remaining norms of sincerity and justification are sufficient to explain the requisite conditions for the uptake of abnormal discourse into normal discourse. To illustrate this point, I considered two agents of sociolinguistic revolution—moral prophets and revolutionary scientists—who are the paradigmatic propagators of abnormal speech in service of social change. I claimed that such forms of agency only seem plausible within practices involving a much more robust constellation of values and norms than the one's posited within Price's account of assertion.[13]

I shall ultimately argue that Price can overcome this challenge by embracing a methodological expansion to include genealogy as a philosophical tool crucial for the subject naturalist.

3.3 Why Subject Naturalists Need Philosophical Genealogy

So far, I have developed the objection that Price's account of assertion fails to account for linguistic revolution. As I mentioned at the outset of this chapter, however, my central claim shall be that subject naturalists like Price ought to expand their methodological toolkits by looking to philosophical genealogy. In anticipation of this line of thought, allow me to provide some initial motivation for this proposal by briefly outlining some of the commonalities between pragmatism and genealogy.

Price's strategy for articulating the norm of truth is to offer a genealogical explanation of why it forms an indispensable part of a broader set of norms belonging to any assertoric practice. Without the truth norm, he thinks, conversation would lack the requisite 'friction' which incentivizes language-users to improve their beliefs (Price 2003, 177). What I find commendable about this approach is its pragmatist orientation. Rather than go in for a theory about the *nature*

[13] These constellations are also much more historically contingent than the norms of assertions which Price recognizes.

of truth, Price wants to understand its value by examining the role that it plays *qua* norm in human life.[14] This is a theoretical orientation that has informed much of Price's writing on the subject, initially finding expression in the claim that philosophy should stick to philosophical explanation rather than (reductionist) analyses of truth (Price 1988, 119).

At the same time, the arguments presented in Section 3.2 suggest that the truth norm and the account of assertion within which it is embedded need to be broadened to include a wider array of norms and values in addition to sincerity, justification, and truth. As it stands, without this broadening the subject naturalist cannot make sense of the phenomenon of sociolinguistic change, especially as it is construed—as Rorty might put it—along the lines of the normalization of abnormal discourse. This point, I believe, is well illustrated by the examples of moral prophets and revolutionary scientists, whose ability to successfully bring about sociolinguistic change requires the availability of practices involving norms *in addition to* sincerity, justification, and truth. This claim resonates with the objections to Price's global expressivism which I outlined at the end of Chapter Two. Cesare Cozzo's conclusion, recall, was that Price's account of truth failed to deliver an explanation of the difference between epistemically virtuous practices and agnostic-persuasive practices; and therefore, failed to account for *our* norm of truth (Cozzo 2012). In a similar vein, one of Henrik Rydenfelt's claims was that the global expressivist had trouble offering a principled account of why certain linguistic practices (namely scientific inquiry) involve a commitment to fallibilism, whereas others do not (Rydenfelt 2019). In part,

[14] In an important respect, Price can be understood as echoing Williams James' famous (and infamous) insight that "*The true is the name of whatever proves itself to be good in the way of belief, and good, too, for definite, assignable reasons"* (James 1955/1907, 59). Although there are passages that give the impression that James was offering an account of the nature of truth (e.g., as an alternative to the correspondence theory or coherence theory), whether or not this was what he meant, what makes him a notable precursor to contemporary writers like Price is his insistence that we need to look at the role that truth plays *qua* value within human life. For instance, James' recognition that we tend to call beliefs true to the extent that they "*help us to get into satisfactory relations with other parts of our experience*" (49), is fundamentally an insight about how the term operates within a dynamic process of inquiry.

the objection that I have developed in Section 3.2 can be seen as a corollary of the issues voiced by Cozzo and Rydenfelt: in certain paradigmatic instances, sociolinguistic revolution depends on social practices instantiating the kinds of norms and values characteristic of epistemically virtuous practices and regional fallibilism—norms and values, that is, which go well beyond sincerity, justification and truth.

Collectively, these criticisms represent two kinds of challenges for Price's account of the truth norm (and to subject naturalism more broadly). First, they point to a kind of *explanatory inadequacy*. Given the parameters of subject naturalism, according to which philosophical inquiry should proceed along the lines of non-representationalist, functional explanations of our linguistic practices, Price develops an account of assertion which, though plausible in certain respects, fails to explain key features of linguistic practice. Second—and this is a point that I shall go on to develop in greater detail in the next chapter—insofar as one is committed to or values epistemically virtuous practices, regional fallibilism, or the possibility of allowing for sociolinguistic revolution, Price's inability to deliver a plausible explanation of such phenomena is *normatively troubling*.

Can Price's account of assertion (in particular) and can subject naturalism (more generally) be adapted to incorporate the additional norms needed to make sense of sociolinguistic revolution, epistemically virtuous practice, and regional fallibilism? In the next chapter I shall argue that they can, and that one promising avenue for doing so is to look to recent work in philosophical genealogy. I will argue that by turning to philosophical genealogy not only can one better diagnose an important methodological shortcoming which gives rise to subject naturalists' explanatory inadequacies, but that doing so also promises a corrective which is compatible with subject naturalism's guiding insights.

To anticipate: we need to recognize that Price's functional explanation of the value of truth is derived from an idealized (and highly abstracted) set of conditions intended to reveal how the truth norm could have emerged in light of generic human interests and concerns. Price explicitly construes this exercise in etiology as a kind of *genealogy* (albeit with very little elaboration of what such a methodological stance ought to involve) and in this respect he prefigures a growing number of contemporary analytic philosophers who have embraced genealogical explanation as an alternative to ahistorical conceptual analysis as a way of gaining insight about philosophical problems. Some of these figures include Bernard Williams, Edward Craig, Miranda Fricker, Ian Hacking, and others.[15]

In the next section, I offer a more detailed discussion of Price's relationship to this body of scholarship through a comparison of his genealogy of the truth norm with Bernard Williams' genealogy of truth and truthfulness (especially as this relates to the second problem for subject naturalism which I identified in Chapter Two). But for now, I want to suggest that some of these more recent writers have identified a component of genealogical explanation which is absent from Price's writing and which, if suitably developed, could yield a broader account of assertion which would be capable of making sense of sociolinguistic revolution.[16] This is a diachronic component by which Price's conception of the truth norm is *de-idealized* through some kind of

[15] Others include Dutilh Novaes (2015), Koopman (2013; 2015), Kusch and McKenna (2020), Queloz (2018a; 2018b; 2019; 2021), Srinivasan (2019). All of these writers claim that genealogical explanation has been around for a long time, notably in the work of the state-of-nature stories proffered by the social contract tradition in liberal political philosophy (e.g., Hobbes, Locke, Rousseau, Marx, Hume) as well as Nietzsche. Foucault is often recognized but not always discussed in detail—with the exception of Koopman (2013). The recent revived interest in genealogy within analytic philosophy presents interesting sociological questions. Why did genealogy become eclipsed in analytic philosophy? One hypothesis is that its reputation became tarnished by positivistic philosophers of science who insisted on a sharp distinction between contexts of justification and discovery. Even defenses of functional explanation (e.g., G. A. Cohen) where couched in the terms of these debates—e.g., Cohen's defense of functional explanation depended on its compatibility with the deductive-nomological model of scientific explanation. In part renewed interest may be due to debates over evolutionary-debunking arguments.

[16] Or what Amie Srinivasan calls "worldmaking" (Srinivasan 2019, 145).

narrative which reveals its increasing functional complexity and historical embeddedness (Kusch and McKenna 2020, 1060).

Although the primary focus of my discussion will be to show that subject naturalists have good reason to adopt the kind of genealogical explanation which Williams develops, it is worth mentioning that pragmatists have also made important contributions to philosophical genealogy which might also serve as models for subject naturalists. Allow me to offer a few examples found in the work of William James and John Dewey.

William James' account of the meaning (or "metaphysics") of paradigmatic moral concepts advanced in "The Moral Philosopher and the Moral Life", can be helpfully understood as a genealogical explanation of the origins of our practices of making moral claims and being subject to obligations (James 1891/1956).[17] Rather than examining empirically the history of Western moral concepts and practices, James' genealogical explanation proceeds by constructing an abstracted and idealized explanatory model on which the function of moral discourse comes into focus. James presents a picture in which creatures with familiar human capacities such as sentience, language use, and the ability to take up the perspective of other people, are gradually added to "an absolutely material world" (James 1891/1956, 189). In doing so, he presents a set of hypotheses about why it is that we employ the moral vocabulary that we do. In Chapter Six, I shall discuss some ways in which James' account bears instructive structural similarities to Price's account of the emergence of the truth norm.

Another pragmatist who employs what can profitably be understood as a genealogical method is John Dewey. In some of his best-known works, Dewey advances functional

[17] This is a claim that I shall develop in Chapter Six.

explanations of the emergence of philosophical concepts and traces their development throughout different historical periods in order to motivate the claim that such concepts are in need of reconstruction. For example, in the first chapter of *The Quest for Certainty*, Dewey aims to explain how modern philosophy came to understand knowledge in the terms that it did—that is, in terms of certainty, privileging theory over practice, and conceiving of knowledge as something timeless and unchanging (Dewey 1929/1984).[18]Ultimately, he suggests that the genesis of these inter-related conceptions of knowledge, and especially the distinction between theory and practice, lies in two possible kinds of responses to the fact that human life is characterized by "precarious possibility" (Dewey 1929/1984, 6). The first kind of response, corresponding to theoretical or intellectual knowledge involves an attempt to "propitiate the powers that environ" by "changing the self through emotion or idea" (3). The second response, corresponding with practical know-how, involves the "invention of arts" aimed at "changing the world through action" (3). In order to appreciate the predicament of modern epistemology, Dewey thinks that we need to examine "the historical grounds for the elevation of knowledge above making and doing" (5). To this end, he offers a brief historical account of the ways in which knowledge has been understood, beginning with Ancient Greek philosophy, through Christianity, and up to our post-Darwinian present.

[18] Dewey's *Reconstruction in Philosophy* exemplifies another attempt to employ a "genetic method of approach" as a "more effective way of undermining… [traditional] philosophic theorizing than any attempt at logical refutation could be" (Dewey 1920/2004, 15). In the first chapter of that book Dewey argues that philosophical theorization emerged in Ancient Greece out of a conflict between two cultural products: on the one hand, the imaginative, emotive material which forms the basis of what Dewey calls "traditional" (i.e., religious and ethical) beliefs; and, on the other, the kind of practical, technical, or "positive" knowledge about nature associated with various crafts. On Dewey's account, Plato and Aristotle responded to this cultural tension by developing "a method of rational investigation and proof which should place the essential elements of traditional belief upon an unshakable basis" (10). This method, along with the metaphysical systems and theoretical dualism which it has produced over the past two Millenia, is, for Dewey, in need of reconstruction given the socio-political conditions of modernity (especially the emergence of the scientific method and liberal democratic political systems).

Whereas both Dewey and James begin by offering explanations of their target conceptual practices in functional, naturalistic terms, Dewey adds an important historicizing, de-idealizing dimension to his account which is absent in the James' (arguably much less ambitious) paper. That being said, the kinds of historical elaborations that Dewey offers in *The Quest for Certainty,* or *Reconstruction in Philosophy* are undeniably course-grained, focusing on broad strokes characterizations of a handful of paradigmatic historical periods rather than more detailed, nuanced discussions of those periods.

For my purposes, drawing attention to these genealogical strands within pragmatist philosophy helps to motivate the claim that Price's subject naturalism—which, after all is a self-consciously pragmatist position—is compatible with philosophical genealogy. This is especially important as there are passages in which Bernard Williams tends to present his position as opposed to pragmatism.[19] In pointing to the resonances between components of Williams' own genealogical framework and elements of James and Dewey's thought, I hope to suggest that the former's opposition to pragmatism is more exaggerated than it needs to be (or is at least more narrow in scope than he tends to present it). At the same time, as I shall go on to argue in the next chapter, Williams' work offers a plausible way of elaborating on and enriching the often-underdeveloped genealogical components of pragmatism. Although my primary task shall be to show that this is the case for Price, I believe that similar lessons could hold for those other pragmatists whom I have been discussing.

[19] For example, Williams writes that, "The pragmatists' claim to have overcome traditional obsessions with the differences between appearance and reality, truth and illusion, and so on, is offered in a setting that not only relies on such ideas itself…but reflects the most elementary Enlightenment optimism associated with them" (Williams 2002, 59-60). In his review of *Truth and Truthfulness*, Rorty expresses his perplexity over Williams' hostility towards pragmatism. He writes, "The similarity between Dewey's and Williams's conceptions of the desirable self-image for heirs of the Enlightenment is, in fact, very great" (Rorty 2002).

CHAPTER 4

OVERCOMING THE STABILIZATION PROBLEM: WHY SUBJECT NATURALISM NEEDS PHILOSOPHICAL GENEALOGY

4.1 Naturalism, Normativity, and the Stabilization Problem

In the previous chapter I argued that in order to allow for sociolinguistic revolution (and ultimately an account of moral and intellectual change), Price's account of assertion needs to be broadened to include a wider set of values and norms, and that one way a global expressivist might do this is to draw from some of the methodological insights of philosophical genealogy. One consequence of this discussion is that these norms of assertion begin to take on an historical vairability. What is involved in "taking one's disagreements to matter" can take various shapes, as the truth norm becomes embedded alongside other values and dispositions within evolving social practices. As Bernard Williams correctly notes, "These dispositions have taken different forms in different historical circumstances" (Williams 20002, 35). This suggests, on the one hand, that there may arise situations in which individuals, or perhaps entire societies come to critically reflect on their commitments to a particular set of norms or are forced to choose between competing configurations in light of other possible alternatives. On the other hand, one might wonder whether the very recognition of the contingency of the normative character of truth might itself call into question people's commitment to it. That is, a possible concern is that reflecting on the fact that one's commitment to finding and telling the truth is not grounded in anything that stands beyond the reaches of time and chance, might itself result in, at the very least, a kind of anxiety about one's own such commitment.

At the beginning of Chapter Three, I claimed that subject naturalism could do more than simply demystify our philosophical perplexities; that it held out the possibility of directing or shaping our attitudes towards their target practices. When it comes to the questions just raised, one might wonder whether Price's subject naturalistic inquiry into functional origins of the truth norm might somehow affect the shape of our commitment to such a norm within current practice. These kinds of questions echo a concern that I raised in Chapter Two—drawing on the work of Paul Redding—according to which, Price's subject naturalism not only fails to secure the normative commitments it seeks to explain, but that it threatens to somehow undermine or subvert them.

In the next section I expand upon Redding's critique in greater detail by looking to its supposed inspiration: the work of Bernard Williams. I shall reconstruct Williams's reasons for thinking that any non-trivial form of naturalistic explanation of our concepts, institutions, practices or values, can potentially undermine our commitment to those items. This problem, I shall refer to as the *stabilization problem*. I will first explain how Williams arrives at an early (albeit incomplete) version of this position and attempt to show that it gains in clarity and plausibility in his later turn to genealogy. Next, in Section 4.3, I distill from Williams's later writings a set of criteria which naturalistic explanations must satisfy if they are to avoid the stabilization problem. In Section 4.4 I shall use these criteria to comparatively assess Williams and Price's genealogies of truth (and truthfulness) as a case study for this problem and its accompanying criteria.[1] Doing so will allow me to better evaluate Redding's criticism of Price sketched in the previous chapter. Finally, I conclude by defending two claims: first, that by paying closer attention to the methodological demands of genealogy, especially its de-idealizing, historicizing component, the subject naturalist can overcome the explanatory inadequacies discussed in the previous chapter. In particular, by looking to empirical history to trace the developments and transformations that the normative character of truth has undergone, subject naturalists will be better able to make sense of the distinction between epistemically virtuous and agonist-persuasive practices, the importance of regional fallibilism, and most importantly sociolinguistic revolution. Second, looking to philosophical genealogy can help subject naturalists avoid the stabilization problem. As we shall see, this is closely tied to Williams' claim that genealogy can, in some cases, perform a vindicatory function.

4.2 Bernard Williams's "notably un-Socratic conclusion"

Bernard Williams is one of the most original and provocative philosophers of the twentieth century. One aspect of his thought for which he is best known is his skepticism about the idea that moral philosophy—conceived as a reflective, theoretical discipline—can provide foundations capable of guiding human action. It is within this broader critique of moral philosophy that his claim that "reflection can destroy knowledge" was initially advanced (Williams 1984, 148). In the previous chapter we saw that a similar worry threatens Price's subject naturalism. That is, at least on Paul Redding's construal, the kinds of naturalistic, genealogical accounts of linguistic practices that Price recommends have the potential to somehow undermine or destabilize commitments to certain concepts, norms, or values. At the

[1] It is surprising that so little attention has been directed to comparing Price and Williams' respective projects. Both share a striking number of commitments.

same time, I suggested that Redding's presentation of this critique remains underdeveloped. In what follows I shall attempt to throw additional light on it.

There are plenty of cases in which the very act of *thinking about* or *reflecting on* some task can thwart a person's ability to carry it out. A pianist might claim to know a sonata by heart; until she begins to think about the movements of her fingers or the notes on the sheet music. An experienced baker could surely be said to know the recipe for sourdough bread which he has made hundreds of times; until, perhaps, he pauses to ask himself whether it requires two or three cups of flour. Williams allows that these and other examples of knowledge-destroying reflection can and often do occur (Williams 1985, 167-8). But his concern in *Ethics and the Limits of Philosophy* is with a fundamentally different set of cases. In so far as the pianist and the baker (along with Williams' own example of the cyclist and the tightrope walker) possess a kind of knowledge that falls prey to reflection, that knowledge is probably best described as practical in nature. In such cases of 'knowing how,' the dismantling power exhibited by reflection strikes us as contingent, in the sense that the practitioner's ability may be easily recovered. Following A. W. Moore, we can say that Williams is, in contrast to these examples, making a claim about *propositional* knowledge that is in some sense, *constitutive* (Moore 2003, 344).[2] There are, in other words, cases in which a person or community's *knowing that* something is the case gets undermined through reflection in such a way that that it cannot be recovered (at least so long as reflection remains a live option for them). Moreover, the idea is not that what *appeared* to be ethical knowledge is revealed, through reflection, to never *really* have been knowledge in the first place. Rather, the idea is that reflection becomes "a part of the practice it considers, and inherently modifies it" (Williams 1985, 168). The modification comes to affect people's ability

[2] See also A. W. Moore (1991, 97-98).

to wield the kinds of concepts that were, prior to reflection, available to them, and whose correct employment was partly constitutive of their prereflective ethical knowledge.[3]

Williams arrives at this claim through his well-known distinction between *thick* and *thin* ethical concepts. Thick ethical concepts such as *lie*, *brutality*, *courage*, *gratitude*, *treachery*, have both a non-evaluative or “world-guided” component as well as an evaluative or “action-guiding” component (Williams 1985, 129-30). By contrast, thin ethical concepts—paradigmatically, *good, right, ought, impermissible*—do not seem to involve a non-evaluative descriptive component. For example, I would seem to be making an evaluative claim both when I call the police’s actions “bad” and when I call them “brutal.” But, at least intuitively, my correctly applying the latter concept requires that certain descriptive features of the situation obtain—namely that the police employed force, and that they acted deliberately. The distinction between thick and thin concepts has many interesting metaethical implications (and is by no means uncontroversial), but for our purposes, the significance of thick ethical concepts has to do with the question of whether they can be expressive of ethical knowledge. This is because, at least in *Ethics and the Limits of Philosophy*, Williams is deeply concerned with whether reflection can destroy knowledge involved in the application of thick ethical concepts. So, we should first ask whether, for instance, a person who claimed that the police acted brutally in their treatment of the protestors, could be said to have knowledge.

[3] Moore offers a helpful gloss on this, suggesting that “The way in which the beliefs are undermined is simply by being shown to be false. But the concept itself is not shown to be false. ('True' and 'false' do not apply to concepts.) Rather, the concept is shown not to play one or more of the roles that it was thought to play. The effect is that people no longer want to think in those terms. But the fact remains that, when people did want to think in those terms, they were able to put the concept to use in making certain true judgments, judgments having the right sort of connection with what made them true to constitute knowledge” (Moore 2003, 352).

One way of answering this question is to take up what Williams calls "the ethnographic stance," which involves "the situation of an observer who has an imaginative understanding of a society's ethical concepts and can understand its life from the inside, but does not share those concepts" (Blackburn and Williams 1986, 203-4). From such a standpoint one might imagine an encounter with a "hypertraditional society" in which language-users employ concepts in such a way that is "maximally homogeneous and minimally given to reflection" (Williams 1985, 142). So long as the inhabitants of this society took care to employ their thick concepts in accordance with public criteria or standards, Williams asks: would they be said to have ethical knowledge? The answer, he thinks, depends on one's view of moral discourse. On the one hand, one might think that "the judgments that members of the society make, imply answers to reflective questions about that practice, questions they have never raised" (146). Such an "objectivist model" would require that there are facts about ethical life that obtain independently of the criteria or standards that any ethical community happens to employ. On the other hand, one might adopt a non-objectivist model, wherein "we shall not be disposed to see the level of reflection as implicitly already there, and we shall not want to say that their judgments have, just as they stand, these [further unexamined] implications" (147). Williams argues that it is only on the second, non-objectivist model of ethical practice that practitioners could plausibly be said to have knowledge (148). This is because the inhabitants of a hypertraditional society would employ ethical concepts according to a relatively stable set of criteria which (owing to the presuppositions of homogeneity and non-reflection) would not themselves be liable to question. Things are very different, however, on the objectivist view which, *ex hypothesi* entails that knowledge requires a much wider reflective stance. On this picture, those participants in a hypertraditional society's "judgments have extensive implications, which they have not considered, at a reflective level, and we have every reason to believe that, when those implications are considered, the traditional use of ethical concepts will be seriously affected" (148). But, as Williams explains,

> if we accept the obvious truth that reflection characteristically disturbs, unseats, or replaces those traditional concepts; and if we agree that, at least as things are, the reflective level is not in a position to give us knowledge we did not have before—then we reach the notably un-Socratic conclusion that, in ethics, reflection can destroy knowledge (148).

Here the picture seems to be that the prereflective judgments using thick ethical concepts contain the potential for knowledge, which can somehow become undermined when a wider reflective position is available, from which the "extensive implications" of those judgments are made apparent.

There does seem to be *something* compelling about this line of thinking. When college students first encounter viewpoints which differ markedly from their own, they often come to regard their previous worldviews as parochial, and consequently, may no longer inhabit them in the same way they once could. Gaining this broader standpoint may undermine a person's ability to employ many of the thick ethical concepts she once did—for instance, *blasphemy* or *infidel*. Or, to take one of Williams' examples, a society that once found a central place for the notion of *chastity* might come to "drift away from" the concept once a particular reflective standpoint becomes available from which commitment to the concept seems untenable.[4] At the same time, as many commentators have pointed out (including Williams himself) there is something rather obscure about all of this. Why, for instance, is it an "obvious truth" that reflection has the power

[4] Below I discuss whether Williams is best thought of as saying that reflection can destroy an *individual's* knowledge, or whether reflection can destroy a knowledge within a moral community (or both).

to unsettle or destroy knowledge? [5] Even if it were obvious, still one might wonder what mechanisms are supposed to account for this fact. What specific features of reflection and its relation to our attachment to certain concepts is supposed to explain the unsettling?

Williams comes closest to answering these questions in his discussion of "a vital asymmetry" between the perspectival natures of perceptual judgments and ethical ones (Williams 1985, 149). A tempting explanation of why reflection can destroy ethical knowledge involves the prevalence of moral disagreement. Recognizing that there are other individuals or societies whose moral responses to certain actions or situations differ from one's own might unsettle one's commitments—leading to, perhaps, an inability to apply one's thick concepts in their usual ways. While this insight is certainly *part* of what Williams is getting at, it is not the entire story. That it cannot be stems from the fact that perceptual judgments are often just as standpoint-laden as ethical ones. It is, after all, a philosophical commonplace that secondary qualities can appear differently to different observers. But, as Williams points out, when we reflect on our perceptual knowledge, we see that "the psychological capacities that underly our perceiving the world in terms of certain secondary qualities have evolved so that the physical world will present itself to us in reliable and useful ways" (150). Taking this stance towards our own ability to perceive the world (including the fact that it is sometimes perspectival) in no way threatens our perceptual knowledge—it does not "unsettle the system" (150). We are able to tell a naturalistic story in which to place our perceptual judgements "in relation to the perceptions of other people and other creatures; and that leaves everything more or less as it was" (150). But when we turn to reflecting on the perspectival nature of moral judgments in search of a "second-

[5] It seems, at least prima facie plausible that a reflecting on the extensive implications of some particular concept, might lead a person to strengthen or deepen their attachment to certain concepts. For critical discussions of Williams' claim that reflection can destroy knowledge see Blackburn (1986), Moore (1992, 2003), Altham (1995).

order account of them" are we able to tell ourselves an analogous explanatory story that also justifies? Williams does not think so. The reason is that any such explanation will presumably have to come from the social sciences. Which means that it will invoke capacities "involved in finding our way around… some social world or other, since it is certain both that human beings cannot live without a culture and that there are many different cultures in which they can live, differing in their local concepts" (150). While reflecting on these capacities *explains* how particular groups of people come to navigate their social worlds, Williams doubts that it could amount to a *justification* for a particular set of ethical concepts. These are "reflective considerations" (usually presupposed to fall within the jurisdiction of modern moral philosophy) tasked with providing a justification for those more local ethical concepts whose use was originally in question (151). The problem is that, "discussions at this reflective level, if they have any ambition of considering all ethical experience…will necessarily use the most general and abstract ethical concepts" (152). But since, as we have seen, these thin concepts do not "display world-guidedness" they cannot be "used to answer questions about the original (i.e., local) thick concepts" (152).[6] As A. W. Moore puts it, this reflective process comes to "undermine the conceptual apparatus required even to think in the relevant terms. The people engaging in the reflection can no longer make judgments of the kind that constitute the knowledge, although they can still have enough of a grasp on judgments of that kind, from without, to see that they constitute knowledge" (Moore 2003, 344).

[6] The analogy with perception will break down at this point because such an explanation will fail to provide a "theory of error" (Williams 1985, 151). That is, an account of why people hold wrong beliefs. Presumably, in the case of perception, such a theory is built into our evolutionary account, but we are unlikely to expect to arrive at the same conclusion when it comes to ethical beliefs.

4.2.1 Problems with Williams's Account

Although the comparison between ethical and perceptual judgments goes some way to clarifying Williams' point, I take it that there is still an air of obscurity involved in his account. As it is formulated in *Ethics and the Limits of Philosophy* the argument that reflection can destroy ethical knowledge suffers from the following problems.

First, Williams seems to run together the theoretical task of explaining our use of certain thick ethical concepts, with the task of deliberating about whether we *ought* to employ those concepts. In his review of *Ethics and the Limits of Philosophy*, Simon Blackburn makes this point quite forcefully by noting a serious ambiguity involved in the "inside/outside" dichotomy implicit in Williams' ethnographic stance.[7] "Going outside," Blackburn explains, could mean "taking up a theoretical, naturalistic stance from which the problem is to give a naturalistic *explanation* of our moral natures. Or, it might be taking up a slightly wider *deliberative* position, in which the question is not 'why am I (or why are people) altruistic, disposed to tit-for-tat propensities (and so on)?' but rather why *should* I be?" (Blackburn 1986, 196). In neither case is it immediately clear why reflection (e.g., coming to see certain human virtues as serving some broader social function) should undermine our commitment to those virtues (196). To use one of Blackburn's favorite examples, consider a referee or an umpire who spends her spare time reading up on the social-psychological underpinnings of athletic competition. As the result of this reflective standpoint, she might come to gain a new perspective on why human beings play sports and even come to see her role in such activities in an unexpectedly new light. However,

[7] This argument is part of a more general attempt to show that Williams has been guilty of overlooking a compelling Humean position which insists on keeping these two tasks (i.e., explanation and deliberation) separate.

Blackburn argues, there is simply no reason to suspect that this new-found reflective position will somehow prevent her from making the correct calls or performing her duties as a referee.[8]

A second problem is that there is considerable ambiguity surrounding the *mechanism* or, perhaps *set of mechanisms* which are supposed to do the unsettling in Williams' account. Is the destructive power of reflection due to a discrepancy between the engaged first-person deliberative stance and the detached third-person perspective?[9] Or is the potential unsettling of ethical knowledge due to the recognition that there is no set of ethical concepts that can ultimately be justified on the basis of reflection?[10] Does reflection somehow introduce new considerations that turn out to be incompatible with previously held beliefs?[11] Or is it some

[8] In his response to Blackburn, Williams admits that he views deliberative and explanatory reflection as inseparable from one another insofar as "good deliberative reflection is guided by a good understanding of how things are, and very general deliberative reflection—on Socrates' question, for instance—will be good only if It is responsive to an understanding at a very general level of who we are and what we are doing" (207). Here he seems to be conceding that *non-deliberative reflection*—of the sort the referee has undertaken—will not tend to destroy one's ability to wield thick concepts. I wonder whether Williams is conceding too much here. It is also worth noting that this disagreement seems closely related to their diverging views about the difference between metaethics and normative ethics. That is, Williams denies that the two can be sharply distinguished, whereas Blackburn strongly insists on their separability.

[9] For instance, Williams writes that "Practical thought is radically first-personal. It must ask and answer the question "what shall *I* do?" Yet under Socratic reflection we seem to be driven to generalize the *I* and even to adopt, from the force of reflection alone, an ethical perspective" (Williams 1985, 21).

[10] At one point, Williams suggests that reflection on human ethical practices will ultimately give rise to the following question: "Granted that human beings need to share a social world, is there anything to be known about their needs and their basic motivations that will show us what this world will best be?" His response: "I doubt that there will turn out to be a very satisfying answer. It is probable that any such considerations will radically underdetermine the ethical option even in a given society… Any ethical life is going to contain restraints on such things as killing, injury, and lying, but those restraints can take very different forms. Again, with respect to the virtues, which is the most natural and promising field for this kind of inquiry, we only have to compare Aristotle's catalogue of the virtues with any that might be produced now to see how pictures of an appropriate human life may differ in spirit and in the actions and institutions they call for. We also have the idea that there are many and various forms of human excellence which will not all fit together into one harmonious whole" (Williams 1985, 153).

[11] Sometimes Williams writes as though reflection can lead people to *replace* old beliefs, whereas at other times he makes it seem as though reflection *adds* new beliefs that are, in some sense, incompatible with previously-held ones. Consider the passage in which Williams seems to state his view most clearly. He writes: "I said that reflection might destroy knowledge, because ethical concepts that were used in a less reflective state might be *driven from use by reflection*, while the more abstract and general ethical thoughts that would probably take their place would not satisfy the conditions of propositional knowledge" (Williams 1985, 167, emphasis added). This ambiguity becomes even more pronounced when one asks whether the destabilization of our ethical knowledge piecemeal or wholesale.

combination of these phenomena that Williams has in mind? Unfortunately, there are passages that seem to suggest each of these interpretations.

A closely related ambiguity has to do with what Williams even means by *reflection*. In a key passage cited above, he seems to claim that ethical knowledge gets destroyed because the perspective from which it is to be reflectively understood will inevitably be that of the *social sciences* (Williams 1985, 150). But Williams doesn't explain what exactly this perspective is supposed to involve, or why it is inevitable. Making and enjoying art involves reflection. So too does the study of human history. Do these reflective activities carry with them the same corrosive power as that of social scientific explanation? Moreover, this ignores questions about distinctions within the social sciences—such as whether we are including qualitative or reflexive social science.[12] A final ambiguity is that at times it is not clear whether Williams' arguments are supposed to apply at the level of individuals or at the level of societies (167-8). Some passages and examples make it seem as though reflection is something that a *person* can undergo and which can unseat their ethical knowledge. At other times, Williams suggests that reflection is something that happens at the level of societies—such that a society more prone to reflection will end up with less ethical knowledge. Again, the plausibility of what is meant by reflection and the

This question emerges in J. E. J. Altham's discussion of how thick and thin concepts are supposed to interact on Williams' account (Altham 1995). One reason to think that Williams has overstated his case is that his argument seems to entail that the ethical knowledge expressed by thick ethical concepts is only threatened one concept at a time. It may be that the modern world characteristically relies more heavily on thin concepts—but it seems patently false to say that it is devoid of them entirely.

[12] The fact that these inquiries are sensitive to the first-person perspective could pose a problem to the extent that Williams' argument presupposes that reflection is at odds with such a perspective. As we shall see, his own shift towards a naturalistic genealogical explanation of truthfulness suggests that there are relevant ways of reflecting on ethical life beyond those of the social sciences.

way it is understood to destroy knowledge will depend on which of these two interpretations is considered.[13]

Given these tensions, it is not clear how well we can apply Williams's arguments to Price's subject naturalism. To take just one example, an obvious worry is that Price is not concerned with thick ethical concepts.[14] Fortunately, in later writings, Williams elaborated these important considerations in greater detail. I want to suggest that his turn to genealogy in the book *Truth and Truthfulness* provides the resources to deal with some of these lingering questions from his earlier writings. In the next section I provide a brief overview of Williams' genealogical project in *Truth and Truthfulness.* From this discussion I shall distill a set of criteria that can be used to assess whether subject naturalism gives rise to the stabilization problem, whereby it runs the risk of undermining or subverting commitments to the very practices it aims to explain. On the basis of these criteria, I shall argue that Price's subject naturalistic account of the truth norm ultimately does fall prey to this problem due to his relative neglect of the de-idealizing, historicizing dimension of philosophical genealogy. At the same time, I shall argue that these important components of genealogy can be found in Williams' later writings. By turning to the history of truthfulness Williams is able to elaborate and extend the idealized values hypothesized within his state of nature model such that they come to be understood in light of our more local concerns and interests. By contrast, the abstract, highly generic needs which Price posits in support of his functional explanation of the truth norm are insufficient—at least by themselves—

[13] One reason that distinguishing between the individual and social in this context is important, is that it will have implications for how we are supposed to understand the kind of "necessity" implied in Williams' claim. When reflection destroys knowledge is this a psychological or sociological fact? Or is it a matter of some strong kind of conceptual or logical necessity? For a brief discussion of this point see Moore (2003, 353).

[14] In particular, Price would probably be suspicious of Williams' distinction between thick and thin ethical concepts on the grounds that it looks an awful lot like another instantiation of the bifurcation thesis.

to make sense of our more local concern with and interest in finding and telling the truth. This I shall claim, returning to my discussion from the previous chapters, is evinced by Price's inability to make sense of certain key practices such as sociolinguistic revolution, epistemically virtuous practice and regional fallibilism. Rather than view this as a fatal blow to Price's pragmatism, however, I believe that the challenge simply invites further elaboration on and expansion of the explanatory resources available to the subject naturalist. These are resources which can be found by looking to philosophical genealogy.

4.3 Williams's Naturalistic Genealogy in *Truth and Truthfulness*

Whereas *Ethics and the Limits of Philosophy* was a book directed primarily at moral philosophers, *Truth and Truthfulness* aims to address a broader cultural tension between the demand for *truthfulness* and the *belief in truth*. The demand for truthfulness leads people to question received theories or institutional claims to truth. This might be thought of as the driving force behind the radical claim that writing history is inherently biased or ideological, to the extent that that it is incapable of producing truth claims (Williams 2002, 1). In contrast to these "deniers" (to use Williams' term) whose demand for *truthfulness* leads them to renounce the very idea of there being something called "the Truth," the "party of common sense," regards such skepticism towards the possibility of truth as fundamentally misguided—perhaps as the result of an inadequate theory about the nature of language or thought.[15] In order to preserve a space in which the demand for truthfulness can coexist with commitment to the possibility of truth, Williams attempts to navigate a *via media* between these two extremes. On the one hand, he agrees with the party of common sense that much of what the deniers say about truth is either

[15] By the "party of common sense" I take Williams to have in mind especially analytic philosophers of language who are unsympathetic to criticisms of terms like 'objectivity,' 'truth,' or 'reason.'

self-defeating or misguided. [16] At the same time, he is sympathetic to the latter's concern for truthfulness and does not want to dismiss their critical claims (for instance about institutional bias in writing history) as entirely devoid of merit.

Although he sees this cultural tension as a contemporary phenomenon, Williams thinks that Nietzsche effectively discovered the problem over a century ago (Williams 2002, 12). The latter saw that "the value of truthfulness embraces the need to find out the truth, to hold on to it, and to tell it—in particular, to oneself. But Nietzsche's own dedication to this value, he saw, immediately raised the question of what it is" (13). It is often uncomfortable to believe the truth. Self-deception is frequently more convenient than critical reflection. And although we can, in many instances, give instrumental reasons for not being deceived, doing so fails to make sense of the unconditional value that finding and telling the truth can play in our lives (14).

Not only does Williams follow Nietzsche in taking up this problem, but he follows the latter in his methodological approach to its solution: genealogy. For Williams, a genealogy is "a narrative that tries to explain a cultural phenomenon by describing a way in which it came about, or could have come about, or might be imagined to have come about" (Williams 2002, 20). In part, such a narrative will need to draw on actual history. However, Williams insists that it will also require the use of "fictional narrative, an imagined developmental story, which helps to explain a concept or value or institution by showing ways it could have come about in a simplified environment and in response to highly generic human interests or capacities" (21).

[16] As Williams puts it, "The desire for truthfulness drives a process of criticism which weakens the assurance that there is any secure or unqualifiedly stateable truth" (1). He goes on to note several problems with this position. First, in order to motivate the claim that there is no truth, the deniers require some kind of story to motivate their denial; which, they have no way of granting authority. Second, if as many deniers claim, all claims to truth can be ultimately reduced to "bare" power relations—then they effectively leave themselves powerless—as their ideas require some kind of legitimacy to have any political currency. Third, they end up rendering conversation impossible (9).

These “state-of-nature” stories exemplify a kind of non-reductionistic form of naturalistic explanation, aiming to account for some important aspect of human life—in this case truthfulness—by gesturing to the function or role that it performs in social practice and how it answers to a set of widely shared human psychological capacities and needs (22).

One might wonder: given that state-of-nature stories are *explicitly fictional* (i.e., they are not intended as descriptions of our hominid ancestors from the Pleistocene), how can they serve a theoretical purpose?[17] Williams identifies three core features of state-of-nature accounts which contribute to the explanatory ends of genealogy.

First, these fictional components of genealogies allow one to provide a functional account of some aspect of human thought or behavior that not everyone would expect it to have. [18] For example, in *Knowledge and the State of Nature*, Edward Craig employs a state-of-nature story to suggest that knowledge attributions could have arisen out of the ubiquitous need to identify good informants (Craig 1990, 11).[19]

Second, a state-of-nature account is functional because the relation between “derived” (i.e., more complex) explanandum and some simpler explanans is “rational” in the sense that, given some “imagined circumstances people with the simpler motivations would welcome… a

[17] Matthieu Queloz suggests “that the State of Nature is most illuminatingly described not as a fiction, but as a *model*, which abstracts as much from past human societies as from present ones. The purpose of this model is not, in the first instance, to identify the *historical* origins of [e.g.,] truthfulness, but to identify the *structural* origins. It serves to localise and bring out the function of the virtues of truth relative to certain contingent facts about human beings and their environment” (Queloz 2018, 6).

[18] Although Williams does not put it this way, that the state-of-nature component of a genealogy can help us identify some *unexpected* or *overlooked* functional aspects of our practices resonates with the idea—most evident in the Michel Foucault’s genealogical and archaeological work—that genealogies are in the business of identifying the underlying historical conditions of possibility of our practices. As Colin Koopman puts it, this Kantian inheritance of philosophical genealogy can be thought of as “a critical interrogation of depth conditions” (Koopman 2013, 33).

[19] Craig’s hypothesis is that the central function of the concept of knowledge, is that it is “used to flag approved sources of information” (Craig 1990, 11).

state of affairs in which the more complex reasons would operate" (Williams 2002, 34). Here, the *locus classicus* would be Hobbes' account of why people (driven by egoistic tendencies and living in a world with scarce resources) would welcome the institution of a commonwealth (Hobbes 1996/1651).[20]

Finally, fictional genealogies, like evolutionary explanations derive "the functional from what is not functional or is functional only at a lower level" (Williams 2002, 34). This is because "A story which offered a collective deliberation as the route to the outcome would presuppose what the story is supposed to explain: the people in the "earlier" situation would have already to appreciate the content of the concepts such as [for instance] justice and property, and their connections with reasons for action" (34).

Together, these three features make the state-of-nature component of a genealogy a powerful explanatory device. They offer insights not only into how our practices (or important aspects of them) might have arisen, but they do so by revealing or uncovering a non-evident functionality performed by those practices. Genealogies gesture towards what those practices *do*, what their *point* is for creatures like us. Moreover, in "deriving the functional from the non-functional" state-of-nature accounts leave room for the idea that their target practices are sustained by non-functional (or non-instrumental) reasons or motivations. As we shall see, this last point is important because it means that *in some cases*, genealogical reflection will be able to reveal our practices (or elements of them) to us in a way that does not undermine or subvert our commitment to them.[21] In Williams's terms, genealogies may be *vindicatory* (36).

[20] See especially chapters 13-18.

[21] This is because some practices resist being understood or embraced in purely functional terms.

Allowing for *both* vindicatory and non-vindicatory genealogical explanations marks an important development in Williams's thinking because it constitutes a case in which reflection on some aspects of human life—including what Williams calls an "ethical system" (Williams 2002, 24)—need not unsettle one's prereflective knowledge. [22] But how is the notion of vindication to be understood? It is helpful to look at two examples that Williams presents. The first is David Hume's genealogical account of justice as an "artificial virtue" offered in Book Three of the *Treatise*. In Williams' terms, Hume begins with an explicitly fictional account in which self-interested people with limited sympathy go from a state in which they *lacked* a concept of and disposition towards justice to a state in which they came to possess such a concept and disposition. "The distinctive idea of Hume's account," Williams writes, "is that when it becomes common knowledge that everyone would benefit from certain practice, those practices arise, and they involve a new kind of reason for action, one that essentially refers to other people's having similar reasons for action" (33). Here we have an example satisfying the above-mentioned criteria. That is, Hume identifies a non-trivial function which dispositions towards justice serve (i.e., allowing for social cooperation), that is given in terms of a non-controversial set of more basic background capabilities and needs. Moreover, from the perspective of those in such imagined circumstances, the more complex reasons for acting justly are reasonable ones (34). This last point is important for Williams because it indicates that someone could accept Hume's genealogical account and "still give justice, its motivations and reasons for action, much the same respect as one did before one encountered the explanation" (36). It is in this sense that Hume's account *vindicates* its subject matter.

[22] Indeed, at one point he says that vindication is the central aim of his genealogy of truthfulness (Williams 2002, 90).

Not all genealogies do this. Nietzsche's genealogical explanation of morality is, according to Williams, an example of a subversive enterprise. There are three reasons for this. First, Nietzsche was attempting to explain a large cluster of moral attitudes and values: a broad cultural phenomenon which does not lend itself well to functional explanation. Second, Nietzsche did not try to explain the rise of "modern morality" on the basis of some more basic set of dispositions or needs, but rather appealed to a set of psychological forces—such as resentment or "baffled self-assertion"—that are equally or more complicated (Williams 2002, 37). Third, and finally, because these more complicated motivations are ones with which most people are unlikely to identify, Nietzsche's account needed to posit "unconscious" mechanisms. But this entails that his explanans operated at the level of individual psychological process whereas his explanandum targets a social phenomenon. In sum, Nietzsche tries to make sense of a moral system "by reference to motivations that people have anyway… which very powerfully resists being understood in such terms" (38).[23]

Williams's official aim, then, is more Humean than Nietzschean in so far as he hopes to offer a genealogy of truthfulness which can vindicate. This will be a naturalistic explanation of why human beings came to *value* concern for the truth. It begins with a state-of-nature story involving creatures with basic human needs and limitations, who use language to communicate, and who engage in activities which require information to be gathered and disseminated. The aim is to "derive within this story *values* connected with these activities" (Williams 2002, 38). In particular, Williams is interested in how two basic virtues of truthfulness—*accuracy* and *sincerity*—could have emerged in connection with these basic communicative practices. Given,

[23] This is not to say that Nietzsche's genealogical explanation of modern morality is inadequate (though that is, of course debatable), rather, it suggests that those who accept his explanation will likely have a difficult time sustaining their commitment to that moral system, or central elements of it.

for example, that the perspectival nature of observation gives rise in an epistemic division of labor among social beings, certain dispositions such as the accurate and sincere reporting of information come to be understood as serving an important social function.

Although this state-of-nature story can help make sense of why truthfulness (i.e., the concern for finding and telling the truth) would be valuable within a basic communication system, it does not explain why such virtues of truthfulness like sincerity or accuracy would come to be valued intrinsically in such a society. In other words, the story provides an explanation of how basic dispositions to accurately and sincerely convey information would have emerged in light of practical problems endemic to a simplified model of communication. What it does not explain is why, *from the perspective of those within the system*, those basic dispositions would be understood as valuable.[24] This is because adhering to them can, in particular cases, conflict with a person's immediate interests. The state of nature story is designed to reveal the functionality of sincerity and accuracy for humans in facilitating communication; but it does not follow that in recognizing this fact, participants of a social practice would have reason to sustain their commitment to those virtues of truthfulness. Williams's claim is that commitment to truthfulness (expressed in sincerity and accuracy) is unlikely to persist if it is only valued instrumentally. This is not to say that truthfulness has no function, but that its function cannot be explained without reference to non-functional values. As Matthieu Queloz puts it, "The state-of-nature model itself reveals that the presence of non-instrumental motivations to be truthful is one of the *functional requirements* on the practices of truthfulness, which means that the functionality of truthfulness cannot be accounted for in purely functional terms. Instrumental

[24] As Williams puts it, "the internal role of truth in the belief-assertion-communication system gets us no further at all on delivering the values of truthfulness, once the questions arise to which truthfulness helps to provide the answer—questions that inevitably arise granted that the participants in the system are people, reflective agents, to whom such questions can occur" (Williams 2002, 85).

motives for being truthful cannot render the practice stable enough to fulfil its function" (Queloz 2018, 13).

Can a genealogy of truthfulness go further than this? Is it capable of explaining the emergence of certain practices, concepts or values in such a way that leaves intact the intrinsic concern that practitioners have for those items? Or will it, like Nietzsche's genealogy of morality, end up by subverting its object? In order to vindicate, for Williams, a genealogy must make sense of how some element of practice can come to be regarded as an intrinsic good. Which means that "its value must make sense…from the inside, so to speak" (Williams 2002, 91). More precisely, *x* can be treated as intrinsically valuable in so far as it meets two sufficient conditions:

(i) *x* is necessary for basic human purposes,
(ii) *x* can be coherently treated as an intrinsic good (which means that it won't be unstable under reflection) (92).

One (possible) aim of genealogy is to vindicate some *x* by "constructing" an account of why *x* is intrinsically valuable. To do this, such an explanation will need to be able to say what that *x* is, as well as the structure of other values and virtues which surround and support it (95). Again, following Queloz, the first condition can be thought of as one of *practical exigency*, which amounts to "a matter of having strong instrumental reasons to acquire something, given the needs and purposes one already has," whereas the second involves *conceptual and affective embeddedness* (Queloz 2018, 9). This means that practitioners "must have the conceptual and emotional resources necessary for them to relate truthfulness to other things they value, such as nobility, or freedom from manipulation, and, crucially, their emotions" (9).

As we have seen, the state-of-nature component of a genealogy can satisfy the first condition, but not the second. Within the idealized, generalized context of Williams' state-of-nature account, we come to see why dispositions to accuracy and sincerity would be necessary for people with a capacity for and an interest in sharing information. What it could not deliver was an account of *why* a commitment to those virtues would be seen as valuable from the reflective standpoint of practitioners. Further elaboration is required in order to make sense of how these virtues of truth could come to matter to participants in a social practice. This is, in part, because the virtues of truthfulness and the constellation of values within which they make sense are bound to be subject to historical inflection. At different times and in various places, truthfulness has meant very different things and has been valued in very different ways by very different people. Despite the diversity, for Williams these *reasons* for valuing truthfulness are seldom (if ever) reducible to functional terms (35). Fictional narratives elucidating the emergence of certain values (i.e., sincerity and accuracy) against a backdrop of basic human needs and concerns can highlight their instrumental value, but further explanatory work is required to make sense of their place within a constellation of evolving attitudes and dispositions. This is where the fictional genealogy needs to give way to history. This is the task that the second half of *Truth and Truthfulness* pursues. There Williams attempts to make real historical sense of the virtues of truthfulness as they are expressed in our practices of inquiring into the distant past, our practices involving self-knowledge, as well as the role of truthfulness within liberal societies.

4.3.1 Advantages of Williams's Turn to Genealogy and the Stabilization Problem

With this overview of Williams' turn to genealogy, I can now state some of the advantages it offers over his position in *Ethics and the Limits of Philosophy*—in particular with respect to clarifying the idea that reflection can destroy knowledge.[25] I shall concentrate on four inter-related commitments, all of which should be considered advantages to pragmatists such as Price. For the purposes of my argument, these commitments are especially important as they allow for Williams and Price's views to be more straightforwardly compared, especially when it comes to the question of whether their naturalistic forms of explanation (which both writers characterize as genealogy) are inevitably subversive.

[1] *Broadened naturalism*. In *Truth and Truthfulness,* not only does Williams explicitly thematize the fact that he is interested in *naturalistic* reflection, but the naturalism he adopts is much broader than the one presented in his earlier writings. Whereas *Ethics and the Limits of Philosophy* restricts, if only by way of implication, naturalistic reflection on human life to the social sciences, *Truth and Truthfulness* presents genealogy as a legitimate form of naturalistic explanation. That this constitutes a more expansive conception of naturalism is obviously implied by the fact that genealogy includes fictional and historical elements.[26]

[2] *Expanded role for reflection.* In *Ethics and the Limits of Philosophy* Williams expressed profound skepticism about the idea that reflection could have anything other than a destabilizing effect on our commitments (at least when it comes to our use of thick ethical concepts). By contrast, in *Truth and Truthfulness* this skeptical position is softened substantially. Whereas in the earlier work, Williams is concerned to show that reflection can destroy knowledge, he comes to accept that some reflective activities can actually vindicate our

[25] My aim in this section is not so much to provide a defense of Williams' genealogical approach, but rather to argue that it can help us glean a set of criteria by which to evaluate naturalistic explanations—especially those which take the form of a genealogy.

[26] Note that Price also uses the term "genealogy" to denote his subject naturalistic explanations of linguistic practice. This is fortunate, as it lends initial plausibility to my claim that Williams criteria are suitable for assessing Price's project. As we shall see, Williams and Price do not necessarily mean the same thing by genealogy—but this realization is productive, since it allows us to identify a way of expanding Price's naturalism.

commitments. As we have seen, the aim of the later book is to show that genealogy can do just that.

[3] *Expanded target of inquiry.* Whereas in *Ethics and the Limits of Philosophy* Williams' argument was directed at cases in which reflection can destroy our ability to wield *thick ethical concepts* in epistemically successful ways, in *Truth and Truthfulness* he broadens the class of phenomena potentially affected by reflection to include seemingly any "cultural phenomena" (Williams 2002, 20) including "concepts, reasons, motivations, or other aspects of human thought and behavior" (34). This is especially important for my purposes, as it means that the account given in the latter work can more plausibly be contrasted with Price's naturalism, which is not directed specifically at the use of thick ethical concepts (though it may include them). This third point is directly related to what is, perhaps, the most important development in Williams' thought.

[4] *Shift away from the epistemic status of doxastic states towards self-understanding.* In *Ethics and the Limits of Philosophy,* Williams was concerned with whether reflection could destroy one's ability to employ thick ethical concepts when making *knowledge claims*. This concern with the epistemic standing of beliefs gives way to a very different concern in *Truth and Truthfulness,* whereby the mechanism by which reflection becomes destructive is no longer understood in terms of undermining epistemic statuses of beliefs, but in terms of undermining self-conceptions or by thwarting *self-understanding*. Genealogies are a form of reflection that have the power to unseat our commitments to various concepts, practices, and values. But Williams's answer to the question of *how* they do this is much clearer than his answer to the question of how reflection can destroy ethical knowledge in his earlier work: in short, his view is that genealogies unsettle when they fail to vindicate. That is, when they fail to place those items

into suitable relations with the rest of practitioners' attitudes and values, or by failing to identify some plausible but unexpected function that those practices serve.[27] David Owen aptly describes this (possible) function of genealogies. In contrast to forms of critique (notably *ideologiekritik*) which aim to undermine the epistemic standing of a set of beliefs, for Owen:

> genealogy aims to elucidate a disjuncture between the ways in which we are intelligible to ourselves with respect to some dimension of our subjectivity or agency, on the one hand, and our cares and commitments, on the other. This disjuncture is not a matter concerning our beliefs but of the relationship between a picture or perspective and our capacity to experience ourselves as subjects or agents in the ways that matter to us (Owen 2002, 222).

This shift has two advantages. On the one hand, in allowing that (genealogical, naturalistic) reflection can vindicate or subvert our commitments as a function of self-understanding or intelligibility, rather than in terms of the epistemic status of our beliefs, Williams is able to sidestep thorny questions about the so-called "genetic fallacy"—the idea that the etiology bears no relevance to a belief's truth or falsity.[28] On the other hand, it renders more perspicuous the mechanisms through which the supposed unsettling or subversion (or vindication) is supposed to come about.

With these advantages in mind, I can now state more precisely the criteria that Williams's later writings offer when it comes to assessing whether a given naturalistic explanation is liable

[27] This is why Williams takes such great care to show that genealogical explanations can make sense of intrinsic value and to illustrate the importance of their relationship to a fictional state of nature.

[28] For a general discussion of whether genealogies commit the genealogical fallacy see Koopman (2013, 62-4). For an argument that they do (when aimed at the epistemic status of beliefs) see Srinivasan (2019). For an argument that they can legitimately affect the space of reasons, see Queloz (2018). The reason that this turn frees Williams from the problem is that the genetic fallacy is a question of whether a belief's origins or causes is relevant to questions about its truth value. But concepts, practices, institutions, and values, etc., do not have truth values. One's commitment to these latter sorts of things *can* be subverted by a story about its origins—for instance, when that story explicates them in such a way that one can no longer accept one's commitment to them, when, in under a new light they begin to chafe at one's self-conception.

to destroy commitments. A naturalistic explanation of some target aspect of human thought or behavior can be said to have reflective stability just in case it meets the following conditions:

> [C1] *Naturalistic-pragmatic condition*: the explanation must make naturalistic sense of the practice and show that it answers to some generic human needs or purposes.
>
> [C2] *Continuity condition*: if the explanation posits contexts of formation, it must be continuous with the current historical context (of justification, or self-understanding).
>
> [C3] *Intrinsic value condition*: the explanation must enable participants to make sense of their commitments "from the inside." To do this, it must present its target in a way that coheres with participants' values, emotions, and self-understanding.

These criteria can help us better understand the different ways in which a naturalistic genealogy can vindicate, subvert, or even misfire (that is, to fail as an explanation). Any such explanation for which the above criteria fail to obtain faces what I shall refer to as *the stabilization problem*. My ultimate aim is to employ these criteria in assessing Price's subject naturalism, and in particular, whether his account of the normative character of truth can satisfy them. But before doing so, I shall briefly discuss some of the ways in which these criteria might be violated.

First, consider cases that fail to meet the naturalistic-pragmatic condition. Any purported explanation that either failed to present its target as continuous with the rest of the natural world or that failed to show how it might have emerged as a response to basic human needs or exigencies, would be a poor explanation. [C1] can be violated in at least two ways. On the one hand, a genealogical explanation might end up shrouding its target practice in obscurity, perhaps by making implausible generalizations or by positing needs or capacities that do not fit with other human capacities or what we know about ourselves as from the perspective of the sciences. G. E. Moore's infamous conclusion that human beings have a faculty of moral intuition which allows them to grasp *sui generis* moral facts, would be an example of this kind of failure to meet the naturalistic condition.[29] On the other hand, a genealogical explanation might fail to make sense of its target in functional terms. This could happen if the explanation misidentified the function that some element of human thought or behavior could be reasonably said to serve, or if the explanation failed to identify such a function in the first place. A good example of this kind of failure would be the (hypothetical) position to which Hume's account of the artificial virtues is, ostensibly, a response. In what is partly an important anticipation of Kant, Hume held that our approval of other people's actions stems from our judgments about their underlying motives.[30] To deem an action virtuous on this view requires taking an agent to have acted from a virtuous motive. When it comes to the question of explaining the origins or emergence of virtuous actions, one tempting response is to appeal to a pervasive desire to be virtuous. While, from our current standpoint this explanation makes sense (i.e., it is not unreasonable to believe that people act generously because they want to be generous), Hume contends that such an explanation founders when it comes to making sense of how people could have come to care about acting virtuously in the first place. He writes,

> the first virtuous motive which bestows merit on any action, can never be a regard to the virtue of that action, but must be some other natural motive or principle. To suppose, that

[29] Moore's *Principia Ethica* is best known for its attack on ethical naturalism (roughly, the idea that moral judgements can be analyzed in terms of the natural sciences [Miller 2003, 10-11]). His line of reasoning, now called the open question argument, is that if we take 'good' to be analytically equivalent to some natural property P, then anyone who wanted to know whether an x that is P is also good would have demonstrated their conceptual confusion about the terms involved. But such questions are intelligible. In fact, it is always an open question whether an x that is P is also good. Moore's point is that if such an identification were to obtain, it would always be unintelligible to ask whether an action that promoted pleasure was good. Therefore, 'good' cannot be semantically equivalent to, or analyzed in terms of, some natural property P (Moore, 1962/1903, 6-17). From this, Moore famously concluded that moral properties are *sui generis* indefinable, only to be accessed by some mysterious faculty of intuition (Moore, 1962/1903, 148). And while few 20th-century philosophers found this to be a satisfying view, Moore's appeal to a mysterious moral faculty can be seen as just one instance of a recurring explanatory failure in the history of philosophy which can be traced back at least as far as the British "moral sense" tradition of the 17th and 18th centuries.

[30] Or, as Hume puts it, "'Tis evident, that when we praise any actions, we regard only the motives that produced them, and consider the actions as signs or indications of certain principles in the mind and temper" (477).

> the mere regard to the virtue of the action, may be the first motive, which produc'd the action, and render'd it virtuous, is to reason in a circle (Hume 1978/1740, 478).

The problem with positing the desire to perform a virtuous action (e.g., to act generously) in this context is that doing so presupposes that one could already identify an act as an instance of generosity. But, if one accepts Hume's initial claim (that praising actions as virtuous depends on positing *virtuous motives*), the very ability to identify virtuous acts *itself* presupposes that people are able to regard each other as having acted from a virtuous or generous motive. Therefore, it simply makes no sense to explain the emergence of people's tendencies towards generosity by appealing to some antecedent *desire* to be generous. Someone who posited this kind of explanation would be "Reasoning in a circle" (Hume 1978/1740, 478). In this case, the would-be genealogist has tried to explain the origins of virtuous motives by appealing to some function (i.e., satisfying our desire to be virtuous) which ends up presupposing the existence of that which it seeks to explain.

A second general way in which naturalistic explanations could misfire is through what Nicholas Smyth has called continuity failure. This arises when one assumes (implausibly) that the original conditions to which a practice is said to be a functional response are still a part of the current conditions in which that practice is sustained. For example, some anthropologists have tried to explain the origins of dietary taboos on the basis of the fact that such prohibitions functioned to protect people from food-born illnesses. While this kind of explanation might make sense of how those taboos could have *arisen*, it fails to explain why they would *persist*, for instance, in a society with technological advances like refrigeration and food safety regulations. In such a case, since the initial function is no longer served, the explanation would need to appeal to some further function—for instance that taboos function to promote social cohesion.

To take the example with which Smyth is primarily concerned, a growing number of contemporary philosophers have attempted to argue on the basis of evidence from evolutionary biology for conclusions about the function that morality serves within contemporary societies. Philosophers like Jesse Prinz and Phillip Kitcher have argued that morality functions to promote social stability by reducing tensions between individuals and groups.[31] Smyth finds this inference, "which moves from the evolutionary function to the contemporary one," to be fundamentally misguided (Smyth 2017, 1129). Any functionalist genealogy will need to demonstrate that the conditions which initially necessitated some functional dimension of practice, are to some extent, still present in later conditions in which functional element is thought to persist (1132).[32] When it comes to drawing inferences about the function of moral practices within our present situation from the conditions under which they were thought to have emerged, Smyth thinks that philosophers have failed to appreciate the fact that the function of some practice is highly sensitive to its context.[33]

[31] In *The Ethical Project,* Kitcher presents a sophisticated position called pragmatic naturalism which aims "to explain how early humans could have acquired the capacity for normative guidance as a response to "recurrent altruism failures", and how the ethical project of elaborating moral codes might have emerged" (Kitcher 2011, 103). Kitcher's book can be read as a kind of pragmatic genealogy which tries to show how ethics evolved as a means of coping with the demands of social existence and as a response to the limitations of what he calls psychological altruism. Unlike Price, Williams, Craig, and others, Kitcher's genealogy does not begin with a fictional model, but rather looks to the evolutionary origins of ethical practice.

[32] Matthieu Queloz has argued that there are at least two strategies for overcoming the problem of continuity failure, both of which can be found in Williams' work. The first is to operate at a "*high level of description*" when constructing a state of nature model, and the second is to identify needs within such a model that have a good claim to being universally shared across societies (Queloz 2018). Doing so allows the model to be treated without historical specificity such that it can serve "as much a model of the present as of a given earlier society" (Queloz 2018). I discuss Queloz's response to this problem below.

[33] In particular, Smyth takes issue with the idea that the assumptions on which the evolutionary genealogies depend still hold in contemporary contexts. One such assumption is that the game-theoretical models on which the evolutionary theorists depend, only work in cases of small groups of humans living in close proximity (Smyth 2017, 1134-5). Another is that they depend on factors such as resource scarcity in order to make sense of how moral dispositions could be adaptive (1136). Indeed, when one looks to contemporary society Smyth argues that "these conditions aren't just erased, they're arguably inverted. As Adam Smith famously predicted, certain modern forms of economic exchange make it the case that selfishness can contribute to group stability by provoking mutually beneficial competition (not cooperation!) and thereby increasing available public goods. This, arguably, has created

Finally, naturalistic explanations encounter the stabilization problem when they attribute or posit commitments that cannot be accounted for from the participants' perspective. A violation of [C3] occurs, for instance, (i) when the target is explained in solely functional terms which conflict with how people understand that item, or (ii) when the functional is explained in terms of the non-functional, but where those non-functional reasons or dispositions are not ones that people could accept. Hume and Nietzsche's writings often have this effect of subverting their readers' commitments by violating these criteria. "It would be a very well-padded Christian", writes Edward Craig, "who could accept Hume's account of the origins of monotheistic belief and continue with faith unabated" (Craig 2007, 183). Of course, the genealogist's purpose may be subversion—in which case they will presumably not be bothered by any of this. It is when genealogies give rise to "genealogical anxiety" (to borrow a term from Amia Srinivasan) as an unintended consequence, that destabilization will be seen as a problem (Srinivasan 2019, 128). These third kinds of failure are especially likely to arise in cases where practices display what Queloz calls *self-effacing functionality*, or "the phenomenon whereby a practice is functional, but we do not engage in it for its functionality, and it is only functional because we do not engage in it for its functionality" (Queloz 2018, 1-2). As we have seen, this is precisely the kind of functionality that characterizes truthfulness.

4.4 Assessing Price's Subject Naturalism

The criteria presented in the previous section enable us to reevaluate a forceful criticism of Price's subject naturalism; namely, that it will ultimately render unstable or subvert the normative commitments that it purports to explain. I can now more precisely state this charge as

correspondingly resilient communities which cannot be seriously disrupted by individual selfishness, since they often feed on that selfishness" (1136).

follows: subject naturalism cannot meet the three criteria of reflective stability. In order to assess the merits of this objection, I shall take up the question of whether Price's genealogical conception of the truth norm can satisfy the three criteria.

Both Williams and Price's accounts of truth plausibly satisfy the *naturalistic/pragmatic condition.* Neither account reveals anything that would lead us to believe that our commitments to truth are somehow mysterious or incoherent. And both appeal to very similar generic human needs to which practices involving the norm of truth can be seen as a rational response. For Williams, this is the need for communication and for sharing information. For Price, similarly, truth-talk can be understood as functionally advantageous in so far as it allows for the pooling of cognitive resources and, therefore, allows us to make sense of a linguistic community coming to improve its beliefs.

How do their accounts fare with respect to the other conditions? Williams devotes considerable effort to ensuring that his genealogy of truthfulness satisfies the *intrinsic value condition*. As we have seen, when it comes to the virtues associated with truthfulness, he raises serious doubts about whether a state-of-nature account can deliver anything more than an instrumental justification for them. It was the introduction of the historical, dynamic component of genealogy which allowed sincerity and accuracy to be understood as intrinsically valuable from the standpoint of practitioners embedded in various historical circumstances. One way to think about this move, as we have already seen, is in terms *de-idealization.* As Kusch and Mckenna put it, this aspect of genealogy begins with the "state-of-nature as its starting point and tracks how the concept [for example,] knowledge would evolve and diversify as the simplifications and distortions of the state-of-nature are removed step by step" (Kusch and McKenna 2020, 1060). This process enabled the values expressive of truthfulness to be recognizable in light of our more local concerns and hence, allowing for the possibility of their vindication (i.e., our commitment to those virtues) by showing how they can fit with our reflective self-image. In the next section, I illustrate in greater detail how Williams' de-idealizing use of history contributes to a vindication by taking up his account of the virtue of sincerity.

There are passages in which Price briefly considers something akin to the intrinsic value condition for reflective stability. In "Truth as Convenient Friction" he asks,

> If truth does play the role I have claimed for it in dialogue, wouldn't the realization that it is a fiction undermine that linguistic practice, by making it the case that we could no longer consistently feel bound by the relevant norms? (Price 2002, 180).

His response to this threat of "dialogical nihilism" is to deny that it is a practical problem, contending that "in practice we find it impossible to stop caring about truth" (Price 2003, 180). This is, presumably, connected to the unrecognizability of sociolinguistic practices which lacked the truth-norm.[34] Just as "The discovery that our biological appetites are not driven by perception of preexisting properties—the properties of being tasty, sexually attractive, or whatever—does not lessen the force of those appetites" so too, the philosophical discovery that our talk of truth does not track some metaphysical property, but merely serves as a "convenient friction" should not weaken our drives towards conversational engagement (180).[35]

[34] In response to the objection that the difficulty in even imagining a community of "merely opinionated asserters" would leave us incapable of even interpreting them as *making assertions* Price writes that, "The difficulty we have in holding on to the idea of such a community stems from our almost irresistible urge to see the situation in terms of our own normative standards. There really is a third norm, we are inclined to think, even if these simple creatures don't know it… But the point of the story is precisely to bring this third norm into sharp relief, and hence I am quite happy to allow challenges to the story on these grounds, which rely on the very conclusion I want to draw. For us, there is a third norm. But why is that so? Where does the third norm come from? What job does it do-what difference does it make to our lives? And what features must it have in order to do this job?" (Price 2003, 172-3).

[35] While there may be something to this analogy, I think that Price is underestimating the magnitude of the impact that the realization that human desires and appetites *are* biological functions has had on various systems of human value and commitment. While he might be right that "the folk" are seldom troubled by the news that philosophical realism is implausible, that many of our sexual desires have a functional, biological explanation *is* and presumably *was* quite significant for those who had, for instance, grown accustomed to viewing it along the lines of temptation or sin.

There are two ways of reading Price's denial of the practical possibility of dialogical nihilism: as a *wholesale* claim about all of our discursive practices or as a more *retail* claim, limited to specific cases in which the truth norm might obtain or fail to obtain. Although I think that Price's account has some plausibility when directed at wholesale dialogical nihilism, his view struggles to provide a vindication of our concern with the truth norm at the retail or case-by-case level. Let me explain why.

When taken as a claim about *wholesale* dialogical nihilism, Price's hypothesis has some plausibility. A world in which conflicting beliefs failed to produce disagreement *tout court,* is, indeed, a difficult world to imagine. At the very least, it is not clear that it is a problem that pragmatists (of all people) should encourage us to worry about. But at the retail level, it is unclear whether Price's account could vindicate a concern for the truth norm *in particular cases*. When it comes to questions about *why* people should take their disagreements to matter, the only explanation that Price has to fall back on is a functional account showing how adhering to that norm would (in the long run) lead to improved beliefs within a community.[36] Although this may be a compelling explanation of why the truth norm may have arisen, it is doubtful that it could do much to sustain commitment to that norm in cases in which that commitment is called into question. Some people may perfectly well understand the collective advantages of commitment to conversational engagement, but find themselves tempted to adopt a free-rider policy that parasitically benefits from a broader commitment. What Price's story lacks—and this is precisely what Williams's aims to offer—is any kind of reassurance for the committed who have begun to

[36] As Price nicely summarizes the ides: "The third norm makes what would otherwise be no-fault disagreements into unstable social situations, whose instability is resolved only by argument and consequent agreement-and it provides an immediate incentive for argument, in that it holds out to the successful arguer the reward consisting in her community's positive evaluation of her dialectical position. If reasoned argument is generally beneficial-beneficial in some long-run sense-then a community of Mo'ans who adopt this practice will tend to prosper, compared to a community who do not" (175).

entertain free-rider thoughts. In part, this is because it does not gesture towards any of the other norms, values, or ideals within which the concern for truth finds expression (and with which people are likely to identify). In other words, Price's genealogy does not account for the ways in which, as Williams puts it, the value of our concern with the normative component(s) of truth "always and necessarily go beyond their function" (Williams 2002, 35).

How about the *continuity condition?* This, recall, was the requirement that the background conditions within which a practice P is said to perform function F must still obtain if the genealogist's claim that P still performs F given current conditions is to be justified. We need to ask: are the conditions in which we find ourselves relevantly similar to those hypothesized in Price or Williams' state-of-nature model? Matthieu Queloz has argued that philosophical genealogists can avoid continuity failures by (i) describing their target practices at a high level of generality; and (ii) by showing how such practices can be thought of as responses to some need (e.g., sharing information, improving collective beliefs) that could plausibly be said to belong to any human society (Queloz 2018). Doing so ensures a kind of default continuity between conditions hypothesized in the state-of-nature and those of the present. This is because, as Queloz explains, "At this level of abstraction, the [state-of-nature] model is no more a model of one point on the timeline of history than any other" (Queloz 2018) and is therefore trivially applicable to our current situation. This is, of course, no guarantee that the genealogist will have provided a *good model*, but that employing a high level of description, and "anchoring" their account in seemingly universal needs, the genealogist can avoid the problem of continuity failure (Queloz 2018).

While I ultimately agree with Queloz that Williams' account is able to overcome the problem of continuity failure, it seems to me that this has more to do with his de-idealizing use of history, rather than through descriptive generality or through anchoring his account in human nature. One reason for this is that, by Williams's own lights, there are certain historical developments—such as the invention of historical time, the elaboration of sincerity as personal authenticity, the formation of the sciences—which cannot be anticipated from the standpoint of the state of nature. The question that the continuity condition raises is whether these developments involve substantive-enough changes in either the conditions in which (at least in Williams' case) the virtues of truthfulness are said to discharge their function, or in the function itself, such that the explanatory link between the state-of-nature model and the historical conditions in question is severed. If Queloz is right, then the sheer level of generality at which the genealogist describes the practice, or the ubiquity of the human needs to which they are said to be a response, is supposed to ensure continuity (Queloz 2018).

It seems to me that there is a tension between Queloz's solution to the demand for continuity and the demand that a genealogy can satisfy the intrinsic value condition. Ensuring the latter, as we have just seen, requires some further elaboration of the target practice (hypothesized in the state-of-nature) so that it comes to make sense in light of more local concerns. But it is hard to see how this de-idealizing move can succeed if it remains at a high level of description. On my view, it would be better to acknowledge that there are often important differences between conditions constitutive of the state of nature and those informing actual historical practices, but to insist that it is the de-idealizing component of a genealogy that can bridge this gap. As Catarina Dutilh Novaes puts it, "the interplay between continuity and change is one of the fundamental aspects of genealogy" (Dutilh Novaes 2015). What such inquiries reveal is a "*superimposition of layers* through processes of re-interpretation of previously existing practices, giving rise to new practices which nevertheless retain traces of their previous instantiations"

(Dutilh Novaes 2015). It is by turning to history to track these developments that Williams is ultimately able to provide a kind of *narrative continuity* between the conditions hypothesized in the state of nature, and those which shape our present engagement with truth and truthfulness.

Given that Price's account of the truth norm lacks an historicizing component, there are grounds for worrying that it violates the continuity condition. Although the generic social needs which his account posits—namely, the need for improving collective beliefs— are arguably still with us, the problem is that the ways in which these social needs are satisfied have diversified, and gained in complexity in ways that clearly depart from the mechanism hypothesized in Price's state-of-nature model. Take, for example, the extent to which hyper-specialization and expertise have come to define knowledge production in many parts of the contemporary world. To the extent that these practices do discharge the function of improving collective cognitive resources, they often do so by strategically disregarding the norm of truth or by severely limiting its application. In order for any specialized epistemic community (e.g., scientific, legal, medical) to operate, it must restrict who counts as advancing truth claims (and thus, those with whom it is worth arguing), or to put it in Price's terms, such a community has to set things up so that the truth norm is *not* engaged 'by default.' But if this is right, then there does seem to be a significant discontinuity between the way in which the truth norm functions within Price's state-of-nature model, and the way in which it functions within present conditions.[37] What the subject naturalist needs is some way to bridge this gap.

[37] According to Price's model, the truth norm functions to improve collective beliefs by ensuring that the any new candidate for collective belief gets 'tested' via reasoned argument against as broad of an evidential basis as possible. In the hyper-specialized conditions of the present, by contrast, the truth norm still performs a similar function, but it does so in a different way. Its applicability is limited so that range of new possible candidates for (expert) knowledge are radically narrowed, rather than broadened.

Here, Price has good reason to look to Williams's genealogy of truthfulness, which attempts to show how the generic, highly abstracted communication system of the state-of-nature has developed into precisely the contemporary system of intense specialization which I have just described. An important part of this story involves the way in which the virtues of truthfulness have come to be associated with the value of freedom—and especially freedom of speech—within the political and legal spheres of liberal societies. On Williams's account, the connection between the virtues of truthfulness and freedom have been given quite different and often competing elaborations, which continue to be expressed in debates about whether or to what extent society should impose regulations on the so-called 'marketplace of ideas' (Williams 2002, 212-219).[38] While a genealogy of these developments is itself unlikely to resolve disagreements created by these competing elaborations, it can highlight their importance and can serve to clarify some of the background values and ideals that are at stake in them. Moreover—and this is my main point—in turning to history and our present situation Williams takes a step towards answering the potential charge of continuity failure by showing how our need for sharing and consuming information has actually evolved alongside the virtues of truthfulness, even when these virtues come to discharge their function in new ways.

Thus, while Price and Williams both offer genealogies of the normative dimensions of truth which arguably satisfies the pragmatic/naturalistic condition, I have been arguing that the

[38] In particular, Williams has in mind debates within the context of United States First Amendment jurisprudence. Part of what is at stake in these debates is whether or to what extent the discovery and transmission of truth is facilitated or hindered by regulating the 'marketplace of ideas.' Williams notes that, in some cases, the notion has been taken literally to entail that a free commercial market on ideas (e.g., news, media, books, public debate, etc.) is the best means for arriving at the truth. Other views conceive of an idealized market whereby "the success of a given idea is measured not by its being bought but by its being accepted (Williams 2002, 214). For my purposes, one especially important point that Williams makes is that the reason that scientific inquiry might be thought to off an "approximation to an idealized market" is because "its actual social structure is in important respects an example of a managed market: it involves such things as an increasingly high entry fee int terms of training, and also, necessarily, a powerful filter against cranks" (217).

former's does not satisfy the remaining criteria needed to confer reflective stability with respect to participants' commitments to practices involving those norms. On the one hand, Price violates the intrinsic value condition (by mobilizing only functional considerations in support of the norm of truth); and on the other hand, he violates the continuity condition (exemplified by the fact that the contemporary epistemic division of labor arguably transforms the relationship between the truth norm and the need for pooling information).

In the remainder of this chapter, I shall argue that Williams's more fully elaborated philosophical genealogy offers resources which can help explain the shortcomings of Price's account which, as I alluded to above, stem from the latter's inattention to the need for de-idealization. First, I shall explain in greater detail how Williams' turn to history represents one way of de-idealizing state-of-nature models which can satisfy the intrinsic value and continuity conditions. Then I suggest how a Pricean account of the truth norm might be de-idealized in a similar way by looking to the ways in which it has been transformed and elaborated within different historical settings. One obvious direction in which to pursue this historical elaboration would be to look towards the more local practices—such as epistemically virtuous practices, regional fallibilism, and sociolinguistic transformation—for which, as we have seen, Price's current account has difficulty accounting.

4.4.1 How Historical De-idealization can Vindicate

One reason that the de-idealizing turn to history is important for a genealogy of truthfulness is that it shows how norms of truth are articulated in and bound up with evolving practices. More generally, insofar as they succeed in placing their target practices into relations with values that make sense from the standpoint of reflective practitioners, genealogies become capable of 'constructing' an account of how those practices could be treated as intrinsically valuable (Williams 2002, 92). This, I have been arguing, forms a crucial component of what allows a genealogy to be vindicatory. While such an account may not be able to convince *everyone* of the value of its target, it *can* strengthen the convictions of the already-committed. And in cases in which their conviction is called into question, this strengthening can be invaluable.[39] There are at least two ways of de-idealizing a genealogical state-of-nature story. The first, which could be called *hypothetical elaboration* aims to anticipate, or work out developments of the practices identified in the state-of-nature, in order to hypothesize how they may have transformed in the direction of our more familiar, local ones. A second strategy, which could be called *historical de-idealization* looks to history to excavate the elaborations and transformations that its target practices have actually taken. In order to illustrate how Williams employs both de-idealizing strategies towards the aim of vindicating the virtues of truthfulness, I shall take up his account of the virtue of sincerity.

As we have already seen, on Williams' state-of-nature model, sincerity—construed as a disposition to assert what one believes—comes to perform an important functional role within a generic and idealized system of communication (Williams 2002, 71). Because it facilitates the spread of information within a perspectival world, it makes sense that the disposition to make sincere assertions should have emerged in human societies. This functional account, however, cannot explain why sincerity has come to be valued by participants within any given social practice. In Williams's terms, the state of nature model fails to make sense of sincerity *from the inside*, and thus, has trouble explaining why such a disposition would persist or remain stable (Williams 2002, 91-92). A philosophical genealogy can advance beyond a bare functional

[39] I am in agreement with Queloz on this point. As he puts it, "genealogy is not an instrument of conversion. But it can promote self-understanding, and thereby strengthen the confidence of those who are, in some measure, already disposed to be truthful" (Queloz 2018b, 11).

account of sincerity by bringing it into connection with other values, needs and ideals with which participants in a social practice are likely to already be committed. If this connection can be established, then sincerity can be regarded as a disposition worth cultivating.[40]

One such connection to which Williams points, and which exemplifies what I am calling an *hypothetical elaboration* of the state of nature, is between sincerity and trust (Williams 2002, 96-7). In so far as sincerity can be understood as trustworthiness in speech, this already (at least to some extent) takes us beyond a purely functional account of its value (at least in some circumstances). This is due, in part, to the fact that so many human relationships about which reflective agents care and which help lend significance to their lives are largely constituted by complex, and often tacit expectations and norms requiring dispositions to sincerity in the form of trustworthiness.[41] While this is certainly not true of all (or even most) situations, that it does play a role in these more limited cases suggests that sincerity will be seen as a virtue worth

[40] One possible objection to this move might be that it seems to just shift the question of functionality (and the associated worry of reflexive stability) away from the target practice or value and onto the broader constellation of supporting values within which the target is supposed to make sense. If the worry is that reflective agents will not be able to make sense of their commitments to, for example, sincerity, simply by reflecting on its instrumental value, then what is stopping them from raising functional questions about the other values (e.g., trustworthiness, nobility, honor, moral equality, etc.) with which sincerity comes to be connected? Although Williams does not address this concern, it seems to me that he may have several possible ways of responding to it. One would be simply to insist that the latter values are ones which can safely be assumed to be regarded as intrinsic. This might be seen as plausible to the extent that the supporting values are ones whose prevalence can be gleaned from history in a more-or-less straightforward way. Another response would be to concede that while taken individually, the latter values may be subject to functional analyses, but that when they are connected together they take on a new character, whose value comes to be understood intrinsically. For instance, although I might be able to entertain free-rider type doubts about whether I should be sincere, or a trustworthy person (when such considerations are presented in functional terms), I might be less likely to do so once I come to see both dispositions as forming an integral part of the social relationships that feed into my self-conception.

[41] *Pace* philosophical theorists who posit something like the ‘normal circumstances of trust’ which supposedly serve as the paradigm for social interaction, Williams insists that the enormous range of possible social contexts and complex kinds of human relationships will involve varying types of trust and expectations involving them, and different degrees to which trust is required (Williams 2002, 111). As he puts it, “In trying to understand Sincerity, however, we cannot simply assume those relations. We need to consider the various kinds of communicative expectations that obtain between people who have different kinds of relations to one another—either in general, or in special situations defined by their roles” (111-112). In many cases, these different forms of human relations, and therefore, the conception of trust which they demand will vary with historical circumstance.

cultivating, even if only to a limited extent. This is simply to say that the constitutive role that sincerity and trust play within, for instance, friendship or parent-child relationships explains why one might expect those dispositions be stable across different cultural contexts.

In another sense, the connection between sincerity and trust does not get us very far because these dispositions have come to be understood in radically different ways throughout history (Williams 2002, 95). Hence the need for *historical de-idealization*. On Williams's illuminating account, since antiquity and up until the modern era sincerity has often been closely associated with notions of honor and nobility (Williams 2002, 115). The motivation to say what one believed and the considerations counting against intentionally deceiving others, often stemmed from fear of "disgrace in one's own eyes, and in the eyes of people whom one respects and who one hopes will respect oneself" (116). This is because the need to deceive others belies one's capacity for self-sufficiency and independence. On this traditional picture, the values and emotions sustaining dispositions towards trustworthiness in speech were embedded in a cultural context characterized by social hierarchies which are no longer with us today. Therefore, while it makes sense that such a conception of sincerity would be valued *given those cultural conditions*, we should not expect to find an identical conception today.

Without denying that there still remain connections between sincerity and honor within present practices, one of Williams's key insights is that a modern understanding of sincerity has evolved away from conceptions of honor (linked to social position) towards the ideal that we (ostensibly) live in a world of moral equals (Williams 2002, 117). One's reasons for speaking openly and without deceit are likely to be understood less in terms of one's fear of appearing dependent on other people, than with a sense that others are owed sincerity from a position of moral equality. This is not to say that the more traditional dimensions of sincerity have disappeared, but rather, "the motivations of fear and shame have to be brought into relation with ideas of what we deserve and can expect from one another, when that is no longer a matter of given hierarchies but of the particular relations in which we socially and personally find ourselves" (117). In part, this shift is due to changing moral and political conceptions such as equality and respect (as well as manipulation or domination). But, Williams also suggests, that our modern conception of sincerity is also tied to more prosaic, logistical features of our lives, such as the fact that "In our world, in which there is much private life and many particular contracts, it is easier to keep a secret without telling lies, and there is a marked difference between the two" (117). Granted that the possibilities for interpersonal relationships within a monarchically governed village are, in many cases, quite different from those within a democratically governed cosmopolitan city, it should not be surprising that they require different forms of sincerity and trust.

One especially important response to these changing social conditions, which Williams traces to the 18th century—has been the invention of the notion of *personal authenticity* as an important elaboration of the virtue of sincerity (Williams 2002, 172). As with the notion of freedom of speech (which I mentioned in the previous section) authenticity has been expressed in multiple, often competing ways, and for this reason Williams rightly regards it with ambivalence. One model of authenticity—whose paradigmatic expression can be found in the autobiographic writings of Jean Jacques Rousseau—amounts to the idea that sincere and spontaneous declarations will unfailingly deliver a clear representation of one's true motivations, revealing something like a 'true self' constituted by a relatively stable and unified set of motives (Williams 2002, 178). A second model which, for Williams, is illustrated in Diderot's *Rameau's Nephew*, construes authenticity not in terms of the hope of sincerely expressing one's true self, but as a kind of social achievement which emerges as people learn to negotiate a balance between their tendencies towards idiosyncratic speech and action, on the one hand, and the expectations of their interlocutors, on the other. In contrast to the Rousseauian thought that the virtue of sincerity can serve as a conduit for expressing one's more-or-less stable beliefs and motives, Diderot's picture takes for granted that "human beings have inconstant mental constitution that needs to be steadied by society and interactions with other people" (Williams 2002, 191).

Both models begin with very different conceptions of the self and entail different understandings of the relationship between sincerity, trust, and social cooperation. At the same time, Williams suggests that they can be viewed as competing responses to a common set of socio-political problems which both writers faced, including the need for "finding a basis for a shared life which will be neither too oppressively coercive (the requirement of freedom) nor dependent on mythical legitimations (the requirement of Enlightenment)" as well as "a personal problem, of stabilizing the self into a form that will indeed fit with these political and social ideas, but which can at the same time create a life that presents itself to a reflective individual as worth living" (Williams 2002, 201). While Williams sees serious problems with Rousseau's picture of authenticity (in part because of its connection to a coercive picture of social organization) and thus, recommends the one thematized in Diderot's philosophical novel, I take it that this recommendation is secondary to his overall aim, which is to *vindicate* the value of sincerity by situating it within a structure of values, ideals and practical concerns which are recognizable within our own historical context. In other words, while Williams thinks that there are good reasons for us to adopt a certain conception of personal authenticity, the primary point of his discussion is to elucidate two culturally specific conceptions of authenticity which his readers are not only likely to understand, but with which they are likely to identify—if only imperfectly.

At this point, someone might take the fact that Williams's genealogy of truthfulness reveals multiple, in some cases incompatible conceptions of the virtues of truthfulness to entail that his project will inevitably be subversive. After all, given the multitude of possible forms that truthfulness could take, why should we feel particularly committed to the one which happens to exist within our historical context, especially when there are plenty of options to choose from? Without denying that it is *possible* to draw such a conclusion from Williams' genealogy, it seems to me that this objection misconstrues Williams's aim which was to vindicate the virtues of truthfulness by showing how they could be valued intrinsically within our own cultural setting. This is not the same as vindicating *a particular conception of truthfulness*. Rather than subverting our commitment to sincerity by showing how it has come to manifest itself in incompatible forms, in showing that these conceptions are themselves the locus of (often intense) cultural debate and discussion, Williams only reinforces the idea that sincerity is a disposition worth caring about. Take, for example, his suggestion that a certain model of authenticity (and thus, associated conceptions of sincerity and trust) plays an important role in the politics of group or national identity (Williams 2002, 201). If he is correct about this, then whatever one's view about identity politics happens to be, to the extent that one *has* a view about such issues, one would seem to be committed to the belief that the values and dispositions of truthfulness which are at stake in debates about identity politics are ones that matter.

Having briefly outlined both de-idealization strategies which Williams (and others) have developed, I can now explain how a subject naturalist might go about making use of them. Indeed, given the striking similarities between the starting points of their accounts, Price may be able to simply help himself to much of the historical component of Williams's project. In particular, the latter's attention to the ways in which the history of truthfulness has been embedded within the history of the sciences—and what Williams calls "the history of intellectual integrity" (Williams 2002, 150)—offers one promising avenue for Price to respond to some of the criticisms which I outlined in Chapter Two: namely that he lacks the resources to make sense of epistemically virtuous practices, and fallibilism. That is, Williams does a significant amount of work to show how to arrive at these practices from an idealized functional account of truthfulness within a state-of-nature model.[42]

In addition to offering a way of connecting the value of truth to epistemically virtuous practices and regional fallibilism, Williams's genealogy offers a way for Price's subject naturalism to develop an account of the relationship between the normative character of truth and sociolinguistic transformation. While sketching the details of such an account is clearly beyond the scope of this chapter, we now have a good idea of at least the direction that such an inquiry would need to take. For example, rather than rest content with Price's functional account of truth talk, the subject naturalist ought to investigate how sociolinguistic transformation has been conceptualized in different historical epochs and to track the ways in which the truth norm has functioned within and evolved in light of those conceptions. Within our own contemporary

[42] More precisely, by adopting Williams' historical elaborations of the role that truthfulness has played in history of the sciences Price could avail himself to a dynamic model that accommodates the historical transformations that the truth norm—as postulated in his own state-of-nature model—has undergone. One such transformation will involve the integration of the norm of truth into a broader constellation of values constitutive of fallibilistic practices or those involving the intellectual virtues. A genealogical account that makes sense of these historical transformations will have two important consequences for Price's account of truth. First, it allows him to overcome the problem of explanatory inadequacy by making sense of the connection between the normative character of truth and important dimensions of linguistic practice (which, as we have seen, his current account fails to explain). Second, by placing the norm of truth in relation to these historically emergent practices, it becomes possible for Price to offer an account of its value that goes beyond functional terms. That is, in so far as reflective agents *care* about epistemically virtuous practices or regional fallibilism, the fact that the norm of truth comes to be seen as playing an integral role in those practices can make sense of why the norm should be regarded as intrinsically valuable.

context, for instance, there do seem to be competing accounts of what social transformation involves which are tied to diverging conceptions of truth (and its normative function). Richard Rorty explicitly contrasts his own pragmatist conception of intellectual and moral transformation—which I alluded to in the previous chapter—with "Realist" or "Platonist" accounts which view intellectual and moral progress along the lines of *finding* or *discovering* the truth, rather than as a creative, imaginative endeavor (Rorty 1998, 205). This historicizing dimension of the subject naturalist's genealogy need not take on the normative ambition of *prescribing* which conception of socio-linguistic transformation to adopt. Rather, part of its aim is simply to elucidate the ways in which the truth norm has taken on different shapes, and served different functions, given these different conceptions.

By looking to the history of how sociolinguistic transformation has been understood, tracking its alternative conceptions, and highlighting the way that the norm of truth—along with other values—are connected with these conceptions, not only might Price be able to overcome the problems of explanatory inadequacy (which I mentioned above in section 3.2), but he would be taking a further step towards offering a vindication for the normative character of truth by illuminating its connection to further values and practices with which people are likely to identify. That is, insofar as it can be assumed that participants within liberal democracies value sociolinguistic transformation—for instance in so far as their self- and cultural conception is bound up with the kinds of practices that make moral prophets and scientific revolutionaries possible—employing an historicizing, de-idealizing narrative which situates the norm of truth within such practices can help make sense of *why* that norm is important, *why* commitment to it makes sense.

4.5 Conclusion

I have argued that Price's account of the truth norm ultimately falls short for two reasons. First, it fails to explain important aspects of linguistic practice, such as the existence of epistemically virtuous practices, regional fallibilism, and sociolinguistic transformation. Second, it fails to secure the kind of reflective stability among those reflective agents to whom it is meant to apply. In order to evade these criticisms, Price—and more importantly, those concerned with elaborating subject naturalism—should embrace a broadened, more dynamic conception of genealogy sensitive to the philosophical importance of history. This would provide him with methodological resources to move beyond solely functional explanations of the emergence of practices, and to elaborate ways in which those practices come to be seen as intrinsically valuable.

The second part of this dissertation applies this expanded version of subject naturalism to a set of debates concerning the concept of moral status. To say that an entity has moral status means that people ought to take its interests into consideration when they deliberate about what to do. Rather than provide an account of the grounds of moral status (e.g., in sentience, or the capacity for reason), my suggestion is that philosophers should instead try to understand the function that the concept of moral status plays within our moral practices. Following my contention that an adequate subject naturalism requires a robust genealogical dimension, my account shall consist of two parts: first, a state-of-nature model which attempts to explain why the concept of moral status would have arisen given a highly generic set of human needs and interests; and second, an historical elaboration of the concept of moral status as it has functioned at different times. Thus, I aim to develop a subject naturalist or global expressivist account of moral status which takes seriously both the functional, pragmatic as well as the de-idealizing, historicizing component of philosophical genealogy.

CHAPTER 5

MORAL STATUS

5.1 Introduction

Suppose that before leaving town, a friend asks you to feed her cat, water her plants, and polish her rock collection while she is away. In agreeing to help, it seems uncontroversial to say that you would thereby take on certain obligations, which—barring some exculpating reasons—would go unfulfilled should your friend return to unpolished rocks, wilted plants, and an emaciated cat. Whenever we talk of obligations, it makes sense to ask *to whom* they are directed. In this scenario, it seems clear that you would have an obligation to your friend, in part, because her interests are at stake. And most of us think that other people's interests matter. Even a psychological egoist (if such a character is even coherent) would readily admit that she has obligations towards *herself*. This suggests that, in general, we can and do have obligations towards *people*—both to others and ourselves.

But what about the cat, the plants, and the rocks? Does it make sense to say that you have obligations towards *them*, in *their* own right, or for *their* own sake?

Many people accept that we have obligations to cats and other non-human animals. Neglecting the cat—causing him unnecessary suffering—would harm *him,* regardless of any anguish that doing so might cause your friend. Some people, though probably far fewer, would maintain that you have an obligation to your friend's plants as well. There is a sense in which a plant is "worse off" when it goes without water. One might even say that your failure to water it would thwart the plant's interests. But unlike your friend or her cat, it is doubtful that a plant can feel pain since it lacks a central nervous system. Indeed, it seems unlikely (though there are some who claim otherwise) that plants are aware or have subjective experiences.[1] Finally, some people might insist that we have moral obligations towards rocks and other inanimate objects. Maybe the rocks are rare, or sacred. Perhaps your friend believes that they are inhabited by benevolent spirits. Barring some sort of commitment to animism, however, it is difficult to see how one could be obliged to a rock.[2] This is because it is difficult to make sense of the idea that rocks can have needs, interests, or anything resembling well-being. Rocks just do not seem to care whether they are polished or pulverized.

One way of putting these questions is to ask whether cats, plants, or rocks are the kinds of things that have *moral status*. To say that an entity has moral status means that it can be the direct object of an obligation; that its interests matter and ought to be taken into consideration when we deliberate about what to do.[3] The concept of moral status has come to occupy an increasingly central place within applied ethics. Many of the philosophers who employ the term aim to discern the properties or characteristics by virtue of which moral agents can be said to have moral obligations towards certain entities. A successful theory of the grounds of moral status would, at least ideally, allow us to determine the scope of our obligations, to decide whose

[1] Michael Marder has suggested that plants are capable of a kind of non-conscious awareness (Marder 2013).

[2] For an overview of recent work on animism see Harvey (2014).

[3] I take the notion of moral status to be synonymous with other notions such as "moral standing", or "moral considerability", or even "moral patienthood." Although these terms are often used interchangeably in the literature, there are exceptions. Allen Buchanan, for example, takes "moral status" to be a comparative notion, and "moral standing" to be a non-comparative one (so that two entities could both have moral standing but differ in the degree of their moral status). Buchanan, however, notes that this distinction is stipulative and constitutes a departure from the standard practice of treating the concepts as co-extensive (Buchanan 2009, 346).

interests ought to be given consideration, or, to put it more colloquially, to figure out who gets to be included within “the circle of moral concern.”

The overarching aim of this chapter and the next is to develop an expressivist framework for thinking about these issues, which draws from the methodological insights developed in Part One. Rather than adumbrating the properties or relations that provide the basis for moral status, this approach begins by inquiring into the function and genesis of the concept, asking: “What role does the notion of moral status play within our practices?” and “how is it that, given the kinds of creatures that we are, we came to employ the concept?” Although several key elements of this proposal shall emerge in this chapter, I shall develop it in detail in Chapter Six. The central aims of the present chapter are mostly preliminary to that analysis and include: (i) an overview of the recent literature on moral status; and (ii) an explanation of the central motivations for an expressivist position.

In section 5.2, I outline some of the debates in applied ethics in which questions of moral status have become salient. Then, in section 5.3, I introduce four families of theories of the grounds of moral status. These views, which I take to constitute the standard or orthodox approaches in the literature, all aim to identify a set of properties or relations which are thought to serve as the basis for an entity’s possessing moral status. I argue in section 5.4, that despite substantive differences, these orthodox approaches all share a set of pre-theoretical commitments, which when taken together form a position, which I shall call *moral individualism*. Next, I consider several writers who reject moral individualism, either in part or in its entirety. Typically, these non-standard approaches emphasize the contextual nature of our moral obligations and deny that the bare possession of certain characteristics is sufficient to generate agent-neutral reasons for action independent of our values and practices. I refer to this family of positions as *moral humanism*. This theoretical divergence between moral individualists and their humanist critics, I claim, represents a longstanding issue within the literature on moral status, and marks one of the central philosophical tensions which an expressivist account can ultimately help resolve. I shall refer to this as the *problem of intractability.*

Section 5.5 I consider a second problem for theories of moral status, which I call *the problem of eliminativism*. Recently, some writers have argued that the very idea of moral status is either useless or confusing; and so, ought to be abandoned. I argue that an expressivist approach provides a compelling response to these concerns, and that its ability to do so represents a further motivation for the project. By giving theoretical primacy to functional and etiological questions, the expressivist view I defend offers a promising direct response to the problem of eliminativism.

5.2 Moral Status in Contemporary Philosophy

The most widely cited characterization of moral status comes from Mary Anne Warren, who writes that,

> To have moral status is to be morally considerable, or to have moral standing. It is to be an entity towards which moral agents have, or can have, moral obligations. If an entity has moral status, then we may not treat it in just any way we please; we are morally obliged to give weight in our deliberations to its needs, interests, or well-being. Furthermore, we are morally obliged to do this not merely because protecting it may benefit ourselves or other persons, but because its needs have moral importance in their own right (Warren 1997, 3).

This passage involves several important ideas. First, it limits discussion of moral status to *entities* as opposed to *action-types* or *institutions*.[4] Second, it suggests that for some entity to have moral

[4] Some writers refer to, for instance, "the moral status of abortion," or "the moral status of slavery." I take it that they mean to interrogate whether those actions or institutions are right or wrong, permissible or impermissible.

status, it is not enough for moral agents to have obligations *involving* it. Rather, agents must have moral obligations *to* that entity, for its own sake. These are sometimes called *direct duties* or *direct obligations*. Third, Warren's definition implies that in order to possess moral status, there must be something about the entity which can generate reasons for treating it in certain ways rather than others.[5] In particular, she suggests that it must be capable of having needs, interests, or wellbeing. By listing these requirements disjunctively, Warren's definition is wider than other definitions of moral status. Compare, for example Warren's definition with those offered by Elizabeth Harman and David DeGrazia, which point to a single capacity—i.e., being harmed or having interests—as essential for possessing moral status:[6]

> A thing has moral status just in case harms to it matter morally... A harm to a being "matters morally" just in case there is a reason not to perform any action that would cause the harm and the reason exists simply in virtue of its being a harm to that thing, and simply in virtue of the badness of the harm for that thing (Harman 2003, 174).
>
> To say that X has moral status is to say that (1) moral agents have obligations regarding X, (2) X has interests, and (3) the obligations are based (at least partly) on X's interests (DeGrazia 2008, 183).

Finally, Warren's definition is silent about the content of the duties or obligations involved in having moral status as well as the possibility that there could be different kinds or degrees of moral status. Other writers build these notions into their definitions. For example, Russell DiSilvestro writes:

[5] Many philosophers argue that whether an entity has moral status depends on its *intrinsic properties* (i.e., those properties which it has independently of its relationships to other entities). For example, Jeff McMahan asserts that "Moral status is based on intrinsic properties possessed by an individual that ground moral reasons for treating that individual in certain ways – reasons that may differ from those deriving solely from the individual's interests" (McMahan 2012). Similarly, Thomas Douglas writes that, "To say that a being has a certain moral status is, on this view, roughly to say that it has whatever intrinsic non-moral properties give rise to certain basic moral protections" (Douglas 2013, 466). I discuss this assumption in greater detail below.

[6] While it may turn out that these notions (i.e., interests, needs, well-being, and the capacity to be harmed) are all analytically related, I believe that it is better to leave open the possibility that they are not. This is one reason for preferring Warren's wider characterization.

> If something has serious moral status, then there is a strong moral presumption against harming it, a strong moral presumption against wronging it, and a strong moral presumption against even speaking ill of it, or "cursing" it in any way. If something has serious moral status, then it is owed respect, indeed owed justice, and there is a standing - reason to benefit it whenever possible (DiSilvestro 2010, 12).

In adding the qualifier "serious", DiSilvestro implies that there may turn out to be different types, levels, or degrees of moral status.[7] Some theorists appeal to the notion of *full moral status* which is meant to capture the idea that there is a certain class of entities to whom all moral agents have a suite of stringent and equally applicable moral obligations.[8]

Most discussions of moral status begin with the assumption, taken to reflect a widespread non-philosophical intuition, that all cognitively non-disabled adult human beings have moral status. Such "paradigmatic" human beings are supposed to have guaranteed claims against moral agents—to be treated in certain ways or owed certain things—which depend on certain features

[7] Whether it makes sense to say that moral status comes in degrees is controversial. Many (probably most) authors maintain, for example, that although moral agents have obligations towards both "normal" adult humans and dogs, they have weightier and more stringent obligations to the former—such that if one had to choose, one would be morally required to prevent harm to the human (all things considered). This is often put in terms of the claim that humans (typically) have higher degrees of moral status than dogs (Jawarska and Tannenbaum 2014, 243; McMahan 2012; Douglas 2013). Elizabeth Harman has argued against this picture, suggesting that the notion of "degrees of moral status" is superfluous. Instead, she explicates the discrepancy between the strength of the obligations owed to a human (versus, say, a cat) in terms of the harm mattering more to the human. According to Harman, "We can explain why [a] person's death matters more, morally, than [a] cat's simply by pointing out that the person's death is worse for him than the cat's death is bad for it (Harman 2003, 180). David DeGrazia has sketched two ways of making sense of the notion that moral status comes in degrees. First, on what he calls the "unequal-interest model" any entity that has moral status is given equal consideration, but differences in the kinds of interests that entities have may entail different kinds of treatment. Second, on the "unequal-consideration model" if an entity A has a higher degree of moral status than entity B, then A's prudentially comparable interests will matter more (i.e., be granted greater weight in decision making) (DeGrazia 2008).

[8] See for, example Warren (1997, 4), Jaworska and Tannenbaum (2018). Common candidates for duties owed to beings possessing full moral status include negative duties against harm, killing, or interference, as well as a strong positive duty to aid. Oscar Horta has questioned the very idea of full moral status. The "maximum status that someone could possibly enjoy," he writes, "would entail that any interest of its possessor or any wanton wish that person could have, no matter how trivial, would count for more than all the interests and preferences of all the other entities existing at any time added together" (Horta 2017, 908). This suggests that there could turn out to be "status monsters" (analogous to Nozick's "utility monsters") whose interests would come to outweigh everyone else's—leading to the absurd conclusion that we would be obligated to do everything possible to satisfy their wishes.

or properties that they possess. As we shall see, the question of what set of features or properties *ground* moral status drives much of the literature on the subject.

While some philosophers and non-philosophers hold that all and only human beings can possess moral status, this has come to be a minority position thanks to an extremely influential argument, called the *argument from species overlap,*[9] which runs as follows:

> P1 If all and only human beings have moral status, then there must be some morally relevant set of properties which is (i) common to all human beings; and, (ii) not possessed by any other creatures.
>
> P2 There is no relevant set of properties which is (i) common to all human beings; and, (ii) not possessed by any other creatures.
>
> C1 Therefore, it is not the case that all and only human beings have moral status.

Those who defend this argument typically begin by pointing out that for any of the observable capacities commonly cited as evidence for human exceptionalism—for instance, language acquisition, the capacity for abstract thought, or creativity—there will inevitably be some humans who do not, and others who will never demonstrate those traits. Moreover, proponents of the argument from species overlap assert that the kinds of non-observable properties—for example, having a soul, possessing human dignity—to which philosophers have traditionally appealed to justify the unique or superior moral status of human beings, are either morally irrelevant or specious in some conceptual or metaphysical way.[10] In particular, these

[9] This argument is sometimes called "the argument from marginal cases." For an overview of different versions and uses of the argument see Horta (2014). See also, Singer (2009), Regan (1986), and Dombrowski (1984). For a critique of this argument, see Anderson (2004), Diamond (1978), Crary (2010), as well as Kagan (2016). Eva Kittay has criticized philosophers for drawing invidious and "epistemically irresponsible" comparisons between non-human animals and humans with cognitive disabilities, as such comparisons are often grounded in empirically indefensible understandings of what cognitive disabilities involve (Kittay 2005).

[10] Peter Singer rejects attempts to ground the full moral status on religious or theological consideration on the basis of their lack of evidence and the "desirability of keeping church and state separate" (Singer 2009, 572). He also rejects claims that human beings possess intrinsic moral worth or "human dignity" as empty rhetoric. For a detailed discussion (and rejection) of the idea that humans possess souls, see McMahan (2002, 7-19).

authors typically insist that membership in the human species is morally irrelevant to how an individual should be treated. Following Liao, I shall refer to this as the "species neutrality requirement" (Liao 2010, 160).[11] Accounts of moral status which violate it run the risk of "speciesism"—a rationally unjustifiable privileging of some group on the basis of its species membership analogous to racism or sexism (Singer 2009; 1974).

In the next section I shall consider in greater detail some of the properties commonly thought to ground moral status. But before doing so, allow me to discuss some of the controversial questions that emerge once one abandons the idea that all and only human beings possess moral status. Ultimately, these are questions that an account of moral status is intended to help answer.

On the one hand, the argument from species overlap undermines the claim that all human beings have a higher level or degree of moral status than that of all non-human animals. This has led philosophers to question whether moral agents have obligations to human zygotes, fetuses, human infants, or even to human beings with severe cognitive disabilities such as dementia.[12] On the other hand, philosophers have increasingly come to accept that non-human animals, especially those displaying sentience, can be the direct object of moral obligations. Indeed, David DeGrazia has suggested that the fact that "traditional morality" denies moral status to non-human animals comes close to a *reductio ad absurdum* of the position (DeGrazia 2008, 189).[13]

[11] As Liao explains, "The Species Neutrality Requirement says that an adequate account of rightholding must provide some criterion for rightholding that in principle does not exclude any species and where the criterion can be assessed through some objective, empirical method" (Liao 2010, 160).

[12] See, for example, McMahan (2002), Tooley (1986), Kittay (2005), Harman (2007), Warren (1997), DiSilvestro (2010), Wasserman et al. (2017).

[13] For helpful general discussions of the moral status of non-human animals see Gruen (2017) as well as DeGrazia (1996).

Recent advancements in biotechnology and computing have motivated debates about the moral status of various entities whose existence, until recently, was thought to be confined to the speculations of science fiction. Several bioethicists have taken up the question of the moral status of so-called *post-humans* or *enhanced humans*: beings whose cognitive, physical, or even moral capacities could be greatly increased through genetic, pharmaceutical, or other technological procedures.[14] Some writers have suggested that the use of enhancement technologies could result in beings with a moral status greater than that of ordinary humans. This raises the concern that the creation of beings with "supra-personal moral status" could jeopardize or comparatively diminish the moral status of cognitively non-disabled adult humans.[15]

So far, the debates to which I have been referring have all involved questions about whether or to what extent certain living, sentient creatures have moral status. Several philosophers have questioned these limits, suggesting that non-sentient or abiotic entities may possess moral status. Recently, a number of writers have raised questions about whether machines displaying intelligence might be said to have moral status.[16] I shall examine these debates in detail in Chapter Seven.

[14] For an overview of the debates concerning human enhancement, see *Human Enhancement* edited by Julian Savulescu and Nick Bostrom (2008).

[15] Francis Fukuyama has this concern in mind when he asks, "If we start transforming ourselves into something superior, what rights will these enhanced creatures claim, and what rights will they possess when compared to those left behind?" (Fukuyama 2009). Allen Buchanan has argued that the very idea of a moral status higher than that of persons is implausible, and that even if such an assumption were granted, it would not follow that the creation of enhanced beings with "higher moral status" would somehow nullify the rights enjoyed by unenhanced humans (Buchanan 2009). That being said, he concedes that such a scenario could generate real conflicts of interest between enhanced and non-enhanced beings, which could incur serious moral costs. In response to Buchanan, Thomas Douglas argues that there could be scenarios in which the creation of supra-persons whose moral status did exceed that of normal persons would constitute a kind of "meta-harm" to normal persons consisting in the harm of having one's immunity to other permissible harm reduced (Douglas 2013, 485). Douglas, however, does not think that the possibility of this meta-harm constitutes a decisive reason against enhancements.

[16] For an insightful overview of some of the recent literature on robot rights, see Gunkel (2018). Robert Sparrow proposes a kind of "Turing triage test", meant to indicate the practical conditions under which we would be said to

Similarly, ethical concerns about altering an entity's moral status have arisen in debates about *human-animal chimeras*—organisms composed of the cellular or genetic material of both humans and non-human animals. While some writers have explored arguments about the supposed unnaturalness of such procedures, or the moral confusion they might engender (Robert and Baylis, 2003), others have suggested that "what is distinctively problematic about chimera research is the possibility that the introduction of human material would enhance an animal's moral status to the level of a normal human adult without respecting the moral obligations entailed by that status" (Streiffer 2019; 2005; see also Koplin 2019; DeGrazia 2019).[17]

Since the early-1970s, environmental philosophers and conservationists have argued for ascribing moral status to non-sentient beings, both living and non-living. A central concern for many of these theorists has been to argue for the intrinsic value of ecosystems, species, or natural objects such as mountains or rivers—the idea being that aspects of the environment have value independently of their usefulness or benefit to human consumption or enjoyment. In his 1972 article, "Should Trees have Standing?—Toward Legal Rights For Natural Objects", Christopher Stone argued that natural objects within the environment such as rivers, mountains and forests should be given certain legal rights, independently of the rights granted to their owners or users

have granted moral status to machines. Sparrow argues that this will occur once we are able to have intuitions about "triage" cases (i.e., in which we must saving one of two entities) involving both humans and machines, which preserve the nature of the dilemma (Sparrow 2004). Mark Coeckelbergh has argued for a relational, social-ecological approach to moral status which could conceivably entail that we have direct obligations towards machines (Coeckelbergh 2010).

[17] Streiffer (2019, 2005) identifies two central questions that human-animal chimera research raises with respect to moral status. On the one hand, one might ask, "under what circumstances is it permissible to perform research in which an animal's moral status is enhanced?". On the other hand, one might ask, "under what circumstances would the introduction of human material actually enhance an animal's moral status?" (Streiffer 2019). He posits two possible answers in response to the second question. First, if moral status is grounded in advanced cognitive capacities, then the creation of chimeras with such capacities might be thought to have a heightened moral status. Second, on an anthropocentric view according to which all human beings enjoy a higher degree of moral status than non-human animals, human-animal chimeras might be thought to have a heightened moral status owing to their being part human (Streiffer 2005, 358). Streiffer contends that these two possibilities pose a substantive ethical concern for human-animal chimerical research.

(Stone 1972). Other strategies for attributing not just legal, but moral status to the environment or aspects thereof include *biocentric holism*—the idea that actions are right or wrong to the extent that they promote the integrity and stability of ecological systems as a whole. Drawing from Aldo Leopold's notion of the land ethic, J. Baird Callicott has defended the view that we have moral obligations to "metaorganismic entities" such as ecosystem and biotic multi-species communities (Callicott 2001, 209). Paul Taylor has argued that we have direct moral obligations towards entities such as mountains, forests, or rivers, as well as natural ecosystems, whose inherent worth is grounded in their being teleological centers of life (Taylor 1981).

A related set of questions include whether non-natural objects or artifacts can possess moral status. Andrew Brennan has argued that while natural objects such as rivers or mountains can have moral status, human-made artifacts do not. This, he contends, is because the former lack a kind of "intrinsic functionality," and are, therefore, capable of taking on an indefinite number of functions or roles. By contrast, an artifact's functionality is more or less determined by its design (Brennan 1984). A related question is whether the fact that an artifact is rare, sacred, or perhaps aesthetically valuable can serve as the basis for its having moral status. Mary Anne Warren appears to endorse this view when she proposes the "transitivity of respect principle" according to which, "moral agents should respect one another's attributions of moral status" (Warren 1997, 170). Elizabeth Harman considers, but ultimately rejects the idea that moral status can be conferred by virtue of the fact that something is loved or worshiped (Harman 2007). For Harman, although there are cases in which it might seem intuitively plausible that one confer moral status to X, simply because others worship or care for X, there is no principled way of ruling out cases in which such attributions appear dubious.[18]

While these questions can be explored independently of one another, attention has been directed increasingly at questions of a more general register. Rather than investigate, for instance, the moral status of fetuses, it is common for theorists to inquire into "the grounds of moral status" *as such*. One reason for this is that debates about moral status have tended to involve comparisons between different kinds of entities. Another is the thought that our reasons for attributing or recognizing moral status ought to be rendered consistent, suggesting the need for a general account of the basis of those reasons.

Since the expressivist proposal that I advance in this chapter and the next shall be concerned with this more general register, it will help to map out the standard or orthodox views concerning the *grounds of moral status*. These accounts aim to identify the properties or relations by virtue of which entities have moral status. After describing these standard views, I shall try to make explicit some of their shared assumptions. Then I shall consider several non-standard approaches to questions of moral status which reject these assumptions.

5.3 The Grounds of Moral Status: An Overview

How do we determine the scope of the beings to whom we have moral obligations? In this section I shall outline some of the answers that have been proposed to these questions. My aim is

[18] Harman's strategy is to present two cases, the first involving a group of people who worship a mountain, and the second involving a group of pro-life advocates who love or deeply care for a fetus. Though Harman thinks that there would be something intuitively wrong about defacing the mountain given the first group's beliefs about and concern for it, the mere fact that the pro-life advocates care about the fetus is insufficient to confer upon it moral status. Thus, those who wish to defend the idea that moral status can be conferred on the basis of the fact that people worship or care about something must offer a principled way of distinguishing between the two cases (Harman 2007).

to indicate what I take to be the most widely held views and to briefly discuss some of their advantages and disadvantages. It is important to emphasize that these are broad categories which may themselves include competing positions. Moreover, for many of the accounts considered, there is room for disagreement over whether the proposed properties are necessary, sufficient, or necessary and sufficient conditions for moral status.

5.3.1 Sophisticated Capacities Accounts

Philosophers have proposed that sophisticated cognitive or emotional capacities serve as the basis for moral status. These include the possession of reason, autonomy, self-awareness, having future-directed aims or projects, or the capacity to use language, just to name a few. Some writers bundle together several sophisticated capacities under the banner of "personhood," which is taken to be the ground for moral status.[19]

Immanuel Kant's account of personhood as the basis of respect is paradigmatic in this regard. Kant thought that human beings are uniquely able to act on the basis of reasons. While other creatures may be capable of various sorts of purposive and even intelligent behavior, their ends or aims are necessarily given to them by their nature. Human beings, by contrast, are endowed with a self-reflective capacity which allows us to represent to ourselves the grounds for our beliefs and motivations, and to subject them to various sorts of normative assessment. That we may set our own ends and act on the basis of them is tantamount to the claim that we are autonomous. Kant argued that there was a necessary connection between our autonomy and our absolute value.[20] Creatures lacking autonomy, he thought, "have only relative worth, as means,

[19] See for example, McMahan (2002, 6), Singer (2009), Savulescu (2008), and Tooley (1986).

[20] Philosophers have interpreted this connection in at least two ways. According to the first interpretation, there is some faculty or capacity called Reason which is intrinsically valuable, and which confers absolute value on those

and are therefore called *things*, whereas rational beings are called *persons* because their nature already marks them out as an end in itself, that is, as something that may not be used merely as a means, and hence so far limits all choice (and is an object of respect)" (Kant 1997 [1785], 37).

In a similar vein, some proponents of contract-based normative ethical theories have suggested that moral status depends on sophisticated capacities such as the recognition of mutual advantage or the ability to form and uphold agreements. If moral obligations fundamentally stem from, or are simply constituted by our agreements with others, this would seem to entail that anyone who lacks the ability to form or understand those agreements fails to have moral status.[21] In particular, some contract-based approaches explicitly hold that the ability to use language or to assess and offer reasons is a necessary condition for having moral status.[22]

Recognizing that their views might be taken to permit egregious forms of cruelty towards those who lack reason or the capacity for language, both Kantians and contract-based theorists have invoked the idea that moral agents can have *indirect duties* involving those who do not themselves have moral status.[23] Kant held that a moral agent ought to refrain from "violent and cruel treatment" of non-human animals given that such behavior "dulls his shared feeling of their

who possess it. On the second "transcendental" interpretation, Kant's point is not that Reason is of absolute value but rather, that the very act of valuing something requires that we are able to regard it as something that anyone could endorse. For a discussion, and compelling defense of the second interpretation see Korsgaard (2018, 138).

[21] In *Morals by Agreement*, David Gauthier writes that, "behavior towards animals is quite straightforwardly utility-maximizing, although it may be affected by particular feelings towards certain animals. In grounding morals in rational choice, we exclude relations with non-human creatures from the sphere of moral constraint" (Gauthier 1987, 285).

[22] See, for example, (Carruthers 1992, chapter 5).

[23] Kant thought that our tendency to attribute moral status to non-human animals was the result of a kind of conceptual confusion. As he puts it, "from all our experience we know of no being other than man that would be capable of obligation (active or passive). Man can therefore have no duty to any beings other than men; and if he thinks he has such duties, it is because of an *amphiboly* in his *concepts of reflection,* and his supposed duty to other beings is only a duty to himself. He is led to this misunderstanding by mistaking his duty *with regard to* other beings for a duty *to* those beings" (Kant 1991 [1797], 237).

pain and so weakens and gradually uproots a natural disposition that is very serviceable to morality in one's relations with other people" (Kant 1991 [1797], 238). Similarly, contract theorists often claim that those who fall outside of the contract can be covered by virtue of their relations with those who fall within it. For instance, someone who lacked the capacity to communicate may be directly covered by the contract through a proxy or advocate.[24]

One advantage of sophisticated capacities approaches is that they do not violate the species neutrality requirement and thus, can avoid an undesirable form of speciesism. There is nothing about being human *per se* which automatically confers moral obligations. Rather we have moral obligations towards certain entities because they possess complex capacities which provide plausible reasons for certain kinds of treatment. For instance, that someone is autonomous counts as reason for not manipulating or deceiving them. In other words, there is a *prima facie* justificatory connection between the sophisticated properties considered above and various kinds of moral obligations.

Despite these advantages, proposals to ground moral status in sophisticated capacities encounter serious problems. First, they are radically under-inclusive, entailing that many human beings (e.g., infants, those with significant cognitive disabilities) and, at least in some cases,

[24] In *What We Owe to Each Other*, T. M. Scanlon suggests something along these lines, claiming that the contractualist's central moral notion of justifiability to others can be extended to sentient beings who "themselves lack the capacity to assess reasons", through a trustee who could represent them (Scanlon 1998, 183). Scanlon's overall view of moral status, however, is complicated for at least two reasons. First, his theory is concerned with just one aspect of morality: namely, what we owe to others (i.e., those capable of holding judgment-sensitive attitudes). Scanlon admits that there "are different kinds of moral values" as well as "different ways of being morally significant" which his book does not address, and that these could conceivably have implications about the scope of our moral obligations. Second, Scanlon seems to endorse what I shall later call a relationalist approach to moral status. In an oft-quoted passage, he writes: "The mere fact that a being is 'of human born' provides a strong reason for according it the same status as other humans. This has sometimes been characterized as a prejudice, called "speciesism." But it is not a prejudice to hold that our relation to these beings gives us reason to accept the requirement that our actions should be justifiable to them. Nor is it prejudice to recognize that this particular reason does not apply to other beings with comparable capacities, whether or not there are other reasons to accept this requirement with regard to them" (Scanlon 1998, 185).

many if not all non-human animals lack moral status.[25] Second, despite the *prima facie* relevance of personhood or the ability to form compacts to our moral lives, it seems arbitrary to suggest that those capacities are *uniquely* suited to form the basis of moral status. That a dog can experience pain seems like a reason for not subjecting him to needless suffering, whether he is sapient seems irrelevant.

There are a few ways of avoiding these objections. First, a sophisticated capacities theorist might appeal to the possibility of different types or levels of moral status, and argue that while, for instance, personhood might be required for "higher" or "full moral status," it is not a necessary condition for a lower level.[26] Alternatively, they might argue that even incomplete realizations of sophisticated capacities is sufficient to ground moral status.[27] A third strategy,

[25] A common complaint against the notion of "indirect duties" (noted above) is that they do not provide agent-neutral reasons for respecting the interests of those who are not able to fall within the contract. As Tom Regan puts it, "it seems reasonably certain that, were we to torture a young child or a retarded elder [sic], we would be doing something that wronged him or her, not something that would be wrong if (and only if) other humans with a sense of justice were upset" (Regan 1986, 182-3). Christine Korsgaard has argued that the Kantian "indirect duties" view towards non-human animals verges on being incoherent. Kant's view combines two separate claims: (i) that our duties to treat animals well are really directed at ourselves and other humans; and, (ii) that such duties stem from the negative consequences that various ways of treating animals might have on our characters or moral emotions (Korsgaard 2018, 101). The problem is that, at least in certain cases, in order to treat animals well, one must take up certain attitudes towards them which, in part, involve taking their interests into account for their own sake (105).

[26] Many of the views considered in the next section adopt this strategy. Jeff McMahan and Peter Singer both advance a two-tiered account of moral status which distinguishes between persons and sentient non-persons. For example, Singer's position is that all sentient creatures have a basic level of moral status which entails that their interests ought to be weighed equally with like interests. However, he acknowledges an important normative difference between, on the one hand, the interest that persons have in continuing to live, and, on the other hand, the interests that non-persons do. He writes that "there is greater significance in killing a being who has plans for the future—who wishes to accomplish things—than there is in killing a being who is incapable of thinking about the future at all but exists either moment to moment or within a very short time horizon… It is, other things being equal, much less a tragedy to kill that sort of being than to kill someone who wants to live long enough to do the sorts of things that humans typically want to achieve over the course of their lives" (Singer 2009, 576). Kagan (2010), Savulescu (2008), and DeGrazia (1996; 2008) hold similar views. The thrust of this idea also seems to be captured in Robert Nozick's dictum: "utilitarianism for animals, Kantianism for people." According to this thought, moral agents ought to "(i) maximize the total happiness of all living beings; (2) place stringent side constraints on what one may do to human beings" (Nozick 1974, 35-42).

[27] In a series of recent papers, Agnieszka Jaworska and Julie Tannenbaum have argued that the incomplete realization of sophisticated capacities is, under the right circumstances, sufficient to confer full moral status (Jaworska and Tannenbaum, 2014; 2015). They begin by granting that full moral status is grounded in the kinds of sophisticated cognitive capacities that "self-standing persons" possess (Jaworska and Tannenbaum 2014, 243).

which I shall discuss in section 5.3.3, argues that moral status can be conferred on the basis of the potential to exercise sophisticated cognitive capacities. This would allow that human infants, for instance, have moral status even though they currently lack the sophisticated capacities in question. Finally, one might identify some less sophisticated capacities which are able to confer moral status to a broader range of beings. Allow me to briefly consider some of these latter views.

5.3.2 Rudimentary Capacities Accounts

Some widespread capacities which could serve as a more inclusive basis for moral status include, but are not limited to, sentience (the ability to experience pleasure or pain), awareness, consciousness, the capacity to care, being alive, being the subject-of-a-life, or having some kind of teleological orientation.

Utilitarians often embrace an account of moral status along these lines.[28] When it comes to deciding whose interests ought to be factored into the utilitarian calculus, Jeremy Bentham famously claimed that, "the question is not, Can they *reason*? nor, Can they *talk*? but, Can they *suffer*?" (Bentham 1789/2000, chap. 17). The most well-known contemporary proponent of

While human infants and adults with severe cognitive disabilities may be unable to fully engage in the kinds of activities that self-standing persons typically do, their capacity to participate in partial or rudimentary forms of such activities is sufficient to justify their having a higher level moral status. This line of thinking stems from the familiar idea that an action's value can depend on its purpose. The authors nicely illustrate this point with the example of two tennis plays who appear to be haphazardly hitting tennis balls around at a tennis court. At a certain descriptive level their actions may appear indistinguishable, whereas at another level, one player may be attempting to learn tennis through practice, whereas the other may be just playing around. Their point is that *because the activities have different ends,* we would be justified in attributing a kind of value to the former activity (i.e., learning through practice) that we would not be justified in attributing to the other.

[28] Wasserman et al. (2017), suggest that act utilitarians do not require a conception of moral status.

this view is Peter Singer, who argues on the basis of the principle of equal consideration of interests, that all sentient life has moral status (Singer 2009; 2011).[29]

Some deontologists ground moral status in rudimentary capacities as well. Tom Regan has claimed that an entity has moral status just in case it is "an experiencing subject of a life" (Regan 1986, 186). In particular, Regan argues that being a such a subject (which involves being conscious, having a welfare, and preferring certain things over others) supports the idea that an entity has inherent value, which in turn guarantees that it holds certain rights (Regan 1986, 187). Christine Korsgaard offers a substantive revision of the Kantian position discussed above by invoking the notion of "having final goods" as grounds for moral status (Korsgaard 2018).[30] Even philosophers who want to move beyond the theoretical divide between consequentialism and deontology have been attracted to a rudimentary capacities view. David DeGrazia, for instance, argues that "having interests" is a necessary condition for having moral status (DeGrazia 1996).

In addition to being able to avoid charges of anthropocentrism and speciesism, grounding moral status in rudimentary capacities promises to be a much more inclusive approach than the

[29] For Singer, "The essence of the principle of equal consideration of interests is that we give equal weight in our moral deliberations to the like interests of all those affected by our actions. This means that if only X and Y would be affected by a possible act, and if X stands to lose more than Y stands to gain, it is better not to do the act. We cannot, if we accept the principle of equal consideration of interests, say that doing the act is better, despite the facts described, because we are more concerned about Y than we are about X. What the principle really amounts to is: an interest is an interest, whoever's interest it may be" (Singer 2011, 20).

[30] Korsgaard distinguished between functional (or evaluative) goods and final goods. The former are generic terms of positive evaluation that we use to indicate that some activities or conditions allow an entity to perform its function well (e.g., water and light are functional goods for plants). By contrast, we say that some end is a final good (for some entity) when "it constitutes or contributes to the well-functioning of an entity who experiences her own functional condition in a valenced way, and pursues her own functional goods through action" (Korsgaard 2018, 22). For Korsgaard, there is a very tight conceptual connection between having a final good and being an animal. In fact, she suggests that "The final good came into the world with animals, for an animal is, pretty much by definition, the kind of thing that has a final good" (21). And because she thinks that "having a final good is the ground of moral standing… it follows that we have no duties to plants and sponges" (23).

first cluster of views considered above. It obviously allows that cognitively non-disabled adult humans, children, and those with significant cognitive disabilities have moral status. It also gets many non-human animals in the door as well, while leaving open the possibility that moral agents could have direct obligations towards intelligent machines.

Yet the appeal to rudimentary capacities encounters several problems. First, unless one distinguishes between different levels or degrees of moral status, these views conflict with the intuition that we have more stringent and perhaps weightier obligations towards human beings than we do to creatures who possess minimal awareness. In a similar vein, when taken as a necessary condition for equal and full moral status, a number of these views carry the unwanted implication that forms of agriculture and pest control are immoral. As Warren puts it, "unless the lives and happiness of beings that are not self-aware are worth little or nothing to them, giving equal consideration to their interests precludes activities essential to human health and survival" (Warren 1997, 82).

A second problem for rudimentary capacities views is that they still fail to deliver criteria for moral status that are adequately inclusive. Sentience, awareness, or consciousness, if construed as necessary conditions, arguably deny moral status to some humans, for instance those who are comatose. These views would also entail that plants, rivers or ecosystems lack moral status, a conclusion that does not sit well with many environmental philosophers.[31]

Whether one accepts that we have direct obligations towards non-sentient or even non-living entities may ultimately depend on one's values and intuitions. For instance, Mark Sagoff

[31] Taken together, these first two problems point to a central challenge that all accounts of the grounds of moral status must face. On the one hand, they want to avoid being too narrow or underinclusive—failing to account for certain entities that do, intuitively, have moral status. On the other hand, if they are too wide or overinclusive, they invite more room for conflicts of obligation which make moral deliberation more cumbersome. For a discussion of this problem, see Jaworska and Tannenbaum (2018).

has argued that the values motivating some forms of environmentalism are fundamentally irreconcilable with those of animal liberation movements (Sagoff 1984). On the other hand, such conflicts may be compounded by disagreements about the application of terms like “consciousness,” “sentience,” and especially “interests.” Recall, for instance, that on DeGrazia’s view, “having interests” is a necessary condition for possessing moral status. There is an ambiguity here involving a normative as opposed to a descriptive sense in which something can have interests. Given a descriptive reading, certain forms of treatment or outcomes can be “in something’s interest” regardless of whether it is sentient. Plants have an interest in being watered, parasites have an interest in finding a host. These claims are often understood in terms of an entity’s relationship towards some functional goal or *telos.* Thus, there seems to be room to argue that one could have obligations towards non-sentient entities, such as plants, rivers or ecosystems—provided that one is prepared to grant some functional sense in which their behavior or responsiveness is somehow goal-oriented. On a normative reading—which seems to be what DeGrazia has in mind—there is a necessary connection between “having an interest” and sentience insofar as “having an interest in X” implies that one is aware of X, and has some kind of evaluative stance towards X (DeGrazia 1996, 40; 2016, 23).[32]

Moreover, a related problem for (some) rudimentary capacities views is that attributions of cognitive capacities or mental states to non-human entities are (at least in some cases) highly controversial.[33] Some philosophers are skeptical about extending attributes such as “beliefs” or “consciousness” beyond human cases. Donald Davidson has argued that attributing beliefs goes

[32] In a recent paper, DeGrazia suggests that without a reasonably worked out theory of well-being, attributions of interests to plants end up being question-begging (DeGrazia 2016, 23).

[33] For an insightful historical overview of the moral status of non-human animals in the history of philosophy, see Gary Steiner’s *Anthropocentrism and its Discontents*.

hand in hand with language use, and therefore cautions against attributing doxastic states to animals (Davidson 1984).[34] For very different reasons, Thomas Nagel insists that qualitative claims about animal consciousness will be unavoidably speculative and anthropomorphizing (Nagel 1974).[35]

Finally, Wasserman et al. (2017) point to a tension which any of the views considered so far must somehow negotiate. Moral status, these authors argue, is commonly understood as both a threshold and a range concept. This means that it picks out beings who surpass some stipulated level or threshold while treating as equal those who surpass the threshold to varying degrees. These dimensions show up in both the categorical and egalitarian forms which many of our moral practices can take. For example, "We do not think that the more highly intelligent, more deeply self-conscious, or more fully autonomous among us have a higher moral status than the rest, even those near the edge" (Wasserman et al. 2017). The problem is that both components try to "impose moral discontinuity over psychological continuous attributes" (Wasserman et al. 2017). Given that capacities such as consciousness, sentience, or rationality seem to be matters

[34] In particular, Davidson argues that "a creature cannot have thoughts unless it is an interpreter of speech of another" (1984, 157). Understanding what a speaker means, presupposes that an interpreter can attribute to them a multitude of beliefs and intentions. At the same time, attributing beliefs and intentions "must go hand in and with the interpretation of speech" (163). This is because, as Davidson explains, "without speech we cannot make the fine distinctions between thoughts that are essential to the explanations we can sometimes confidently apply. Our manner of attributing attitudes ensures that all the expressive power of language can be used to make such distinctions. One can believe that Scott is not the author or *Waverly* while not doubting that Scott is Scott; one can want to be the discoverer of a creature with a heart without wanting to be the discoverer of a creature with a kidney… The intensionality we make so much of in the attribution of thoughts is very hard to make much of when speech is not present" (163).

[35] In order to argue that the subjective character of conscious experience—i.e., the idea that "an organism has conscious mental states if and only if there is something that it is like to be that organisms" (436)—cannot be explained on the basis of any reductionist account of the mind (i.e., physicalism, behaviorism), Nagel denied that any amount of physiological knowledge, behavioral observation, introspection, or empathy would ever enable us to grasp the subjective character of the consciousness of another being like a bat or a wasp.

of degrees, an adequate account of moral status must explain why those differences above the threshold should be discounted while the threshold itself should remain salient.

The following sets of views attempt to correct for many of the objections leveled against the sophisticated and rudimentary capacities approaches.

5.3.3 Potentiality Accounts

Some theorists argue that moral status is grounded not only in occurrent or actualized capacities, but also in an entity's *potential* to exercise them. A major attraction of this view is that it promises to allow that all human beings have full moral status while avoiding the charge of speciesism.

Russell DiSilvestro offers an elaborate book-length defense of a potentiality account of moral status. For DiSilvestro, all human beings either possess or have the potential to possess a set of "typical human capacities" (e.g., thought, creativity) which, in turn, is sufficient for them to possess "serious moral status" (DiSilvestro 2010). On this view, even a human zygote with congenital conditions precluding it from developing thought or awareness would have serious moral status due to its "higher-order capacity" (i.e., capacity to have the capacity) for typical human capacities. Moreover, their failure to possess such higher-order capacities entails that non-human animals lack serious moral status on this view.[36]

Shelly Kagan has proposed a view called *modal personism*, according to which the modal properties of creatures who could have become persons, are sufficient for their having moral

[36] The consistency DiSilvestro's position depends on the remarkable claim that it is metaphysically possible that presently-unavailable technologies could be developed which would *potentially* correct the human zygote's congenital conditions, but that it is metaphysically impossible for future technologies to confer "typical human capacities" such as thought or rationality to non-human animals (DiSilvestro, 57-58, 151-152).

status (Kagan 2016, 16). Like many others, Kagan allows that both persons and sentient non-persons have moral status, even if the former do so to a greater degree (Kagan 2016, 9). But by insisting on the moral relevance of modal personhood, Kagan contends that even human beings who are not persons have moral status. Moreover, he maintains that this account avoids the charge of speciesism. As he explains,

> membership in the species is not, in and of itself, the morally relevant feature. What really matters is the modal property itself — the fact about what the individual could have been. And in particular… what membership in a person species reveals is that even an individual who is not in fact a person nonetheless could have been a person (16).

S. Matthew Liao has recently argued that "the genetic basis for moral agency" is a sufficient condition for possessing moral status (Liao 2010). Like DiSilvestro and Kagan, Liao aims to support the claim that all human beings possess full and equal moral status. On his view, this follows from the fact that every human being will arguably possess "the set of physical codes that generate moral agency" located in their genome (Liao 2010, 164). At the same time, Liao claims that his position avoids the charge of speciesism, as it identifies a sufficient, but not necessary condition for moral status with a creature's genetic code (164).

Attempts to ground moral status *via* potentiality face serious challenges.[37] The first, which one could call *the conceptual problem*, is that the expression "x has the potential to F," or "y is a potential G," are ambiguous. When construed broadly enough, anything has the potential to be or to do anything. Given sufficient technological advancements, a random lump of matter *has the potential* to be transformed into a normal adult human. Does this mean that all lumps of matter have the potential to become humans? There are obviously different kinds of potentiality, ranging from what is physically or causally possible, to what could occur without logical

[37] These first two challenges to potentiality accounts are discussed by Joel Feinberg (1986).

contradiction. Even on a modest construal, which limits potentiality to what is possible given current technological achievements, grounding moral status on the basis of the potential for certain capacities has unacceptable implications.[38] For example, given the current technological possibility of somatic cell nuclear transfer cloning, every (nucleus-bearing) human cell—for instance, a skin cell—arguably has the same "potential" to become full-fledged human beings as do human embryos (Charo 2001).[39] But if this is right, then potentiality accounts of moral status would seem to be committed to the claim that skin cells have moral status. [40]

[38] Michael Tooley employs the following thought experiment to argue against the claim that human fetuses have right to life by virtue of their being potential persons. Imagine, that by injecting kittens with a special chemical, they could develop the cognitive capacities of a normal adult human, thereby causing them to become persons. Tooley argues that it would be neither morally impermissible to refrain from injecting a kitten with the chemical, nor to interfere with the process in a case in which the kitten was mistakenly injected. From this he concludes by analogy, that "if it is not seriously wrong to destroy an injected kitten which will naturally develop the properties that bestow a right to life, neither can it be seriously wrong to destroy a member of *Homo sapiens* which lacks such properties, but will naturally come to have them. The potentialities are the same in both cases" (Tooley 1986, 81).

[39] See DiSilvestro (2006) for a response. DiSilvestro argues that there is a substantive metaphysical difference between *development* and *generation* which accounts for the difference between the kind of potentiality a human embryo has and the kind of potentiality a skin cell has. For *x* to develop into *y*, he argues, *x* and *y* must retain numerical identity and not undergo a change in kind. Whereas generation does not require that either these conditions be met (Disilvestro 2006, 149). From this distinction, he argues that an embryo can develop into a human, but a skin cell cannot without undergoing a change in kind and numerical identity. This, in turn, allows him to claim that skin cells do not have the same potential to become persons that fetuses do.

[40] Strictly speaking, neither Kagan nor Liao appeal to the notion of potentiality but rather to the notions of "having the modal properties of personhood" and "having the genetic basis for moral agency" respectively. It is, however, hard to resist the conclusion that their views encounter a similar problem. In Kagan's case, depending on how widely one construes the scope of the modal notions, it seems at least possible that non-human animals have the same—or relevantly similar—modal properties as the human cases that he has in mind. For Liao, given that one's genetic basis for moral agency depends on the interaction between one's genome and one's environment, it seems at least conceivable that, given the right kinds of conditions, a non-human animal's genome could be altered in such a way that it serves as the basis for moral agency. As David DeGrazia notes, "One's genome is determined, at a first approximation, by the genetic endowment with which one is conceived and, at a second approximation, by one's original genome plus the effects on it of any spontaneous mutations that accrue over time as one ages. With gene therapy partly in hand and genetic enhancement visible on the horizon, we should recognise that one's genetic constitution can change significantly" (2015, 24). Proponents of the views considered in this section can almost always be expected to respond to this point by making the essentialist claim that such genetic alterations would be identity-destroying. I agree with DeGrazia that these responses are unconvincing given that "no plausible account of numerical identity implies that a nonperson cannot transform into (as opposed to being replaced by) a person. An account that did imply this would imply that you, the person, did not exist as a newborn — who lacked the capacities that constitute personhood — a truly absurd implication. You and the earlier newborn share a single biological life; and the 'two' of you share the same basic capacity for consciousness, making you the same sentient being. Similarly, an ordinary dog who lived a portion of her life before being genetically enhanced could become a person, continuing both the biological life and the sentient life of the pre-enhanced dog" (25).

A second problem facing potentiality accounts involves what Joel Feinberg calls "the logical point about potentiality" (Feinberg 1986). Those who appeal to potential for certain capacities as the grounds for moral status infer illicitly something actual (e.g., a set of rights, being the object of direct duties) from the potential for holding those rights. To appreciate the problem with these types of inferences consider that potentiality accounts seem to license the following claim:

[1] if something is a potential X, then that thing has the rights normally granted to Xs.

But if [1] is true, then it follows that I have the right to command the US military, since I have the potential to become the US president. But this is hardly credible.

Finally, a third problem is that it is not clear that potentiality, genetic dispositions, or modal properties are even relevant to whether some entity has moral status. David DeGrazia makes this point quite forcefully with the following example. Suppose there were born two anencephalic infants, one whose "neural anomaly is due to a defect that originates in utero and not as a consequence of genetic endowment. The other infant similarly lacks the capacity for consciousness, but his deficit is due to genetic endowment" (DeGrazia 2016, 24). According to the views considered in this section, the first infant would have moral status, whereas the second would not. This, however, seems intuitively implausible, and suggests that potentiality or modality are no more relevant to moral status than is species membership.

5.3.4 Relational Views

Some philosophers have argued that moral status can be grounded in the biological or social relationships that an entity bears to others. On a "species-norm account" of moral status, "an individual's moral status depends not on the properties and abilities she actually possesses, but depends instead on the properties and abilities normal for her species" (Wilson 2005, 2). Like potentiality accounts, this would allow all human beings to have moral status, given their relationship to the biological norms of the species. It also avoids anthropocentrism insofar as it is possible that non-human species normally possess characteristics that are sufficient to ground moral status.

Jeff McMahan rejects species-norm accounts because of their unintuitive implications. He considers the case of the Superchimp, a genetically-modified chimpanzee who comes to have the normal cognitive and emotional capacities of a ten year old human.[41] If the species norm account is correct, then given that the Superchimp still belongs to a species which does not normally develop the capacities for personhood, it follows that the Superchimp would lack the higher level of moral status which persons normally have. But this seems hard to accept. Moreover, one might imagine a scenario in which the Superchimp, having spent some time exercising these elevated capacities, were to suffer brain damage, reducing him to the same level of cognitive functioning of ordinary chimpanzees (McMahan 2002, 147). According to the species-norm account, if the same misfortune were to befall a ten-year-old human, the latter would retain a higher level of moral status than the Superchimp; which seems unintuitive.[42]

Eva Kittay has argued that moral status can be conferred through special social relationships including familial and care-giving relations (Kittay 2005, 107, 124). On this view, for example, parents have obligations to care and nurture their children simply by virtue of their

[41] Although McMahan employs this thought experiment to illustrate a point about the concept of fortune (i.e., a way of assessing the good of a life as a whole [McMahan 2002, 146]), as Wilson (2005) notes, the point applies just as well to species-norm accounts of moral status.

[42] McMahan's point is that the species-norm account would have to say that there was nothing unfortunate about the Superchimp's loss of cognitive capacities but that there *was* something unfortunate about the child's. An implication that he takes to be arbitrary: "If the human being and the Superchimp have both fallen from the same height to the same lower state, it seems that either both are unfortunate or neither is" (McMahan 2002, 148).

relationship to them. Not only are these obligations best understood as direct obligations (i.e., they are directed at the child *for its own sake*), they need not depend on intrinsic properties of the child. If cogent, this account would justify a more expansive conception of moral status than that offered by accounts which ground moral status on an individual's sophisticated or even more rudimentary capacities. In particular, Kittay argues that her view supports the claim that cognitively disabled humans—who are involved in social and biological relationships with other human beings—have moral status.

The standard objection to relational views such as Kittay's is that they fail to generate impartial (or "agent-neutral") reasons; and are, therefore, not genuinely accounts of moral status (McMahan 2002; 2005; Jaworska and Tannenbaum 2018). This, of course, is not to deny that biosocial relationships do constitute an indispensable part of moral practice and in many cases are an important source of moral obligations.[43] The problem is that moral agents who do not stand in the relevant relationships to moral patients would not have moral obligations towards them. As McMahan puts it,

> [T]he claim that radically cognitively impaired human beings are specially related to all other human beings would not give Martians a moral reason to treat these human beings any differently from animals, except perhaps an indirect reason deriving from their reason to respect those human persons to whom the impaired human beings would be specially related. Martians might, that is, be morally required to accord the radically cognitively impaired special treatment for much the same reason they would be required to give special treatment to people's pets. Otherwise it would be permissible for them to

[43] According to Jeff McMahan, relationships can have both intrinsic and instrumental moral significance. While some relations (e.g., parent-child relationships) have intrinsic moral significance, other relations (e.g., species relations) do not. One might think that even if species relations do not have intrinsic moral value, they might turn out to have an important instrumental value, for example, by giving rise to a kind of partiality which results in the protection of those who lack the typical moral status-conferring properties. McMahan argues that this way of thinking actually has pernicious effects when it comes to our treatment of non-human animals. He writes, "Just as the darker side of national solidarity is a tendency to denigrate or even dehumanize the members of certain other national groups, so the other side of species partiality is a tendency to treat the interests of animals as morally insignificant. If we compare the number of radically cognitively impaired human beings who benefit from our partiality with the number of animals who suffer from our tendency to regard them primarily as means to our ends, it is hard to believe that the effects of species partiality are desirable overall from an impartial point of view" (361).

> treat the radically cognitively impaired in the ways in which we treat animals, assuming that our treatment of animals is consistent with what is demanded by respect for their intrinsic natures (McMahan 2005, 360).

In response, Kittay has argued that special relationships can generate obligations on the part of those who are not directly involved in those relationships. For example, if (as McMahan seems to grant) parent-child relations can legitimately confer special moral obligations, it would seem that in order for parents to fulfill their obligations, other members of the social body would necessarily need to take on certain obligations to the child (Kittay 2005, 623).

5.4 Moral Individualism and Non-Standard Approaches to Moral Status

5.4.1 Moral Individualism

So far, I have sketched four families of theories concerning the grounds of moral status and mentioned some of their advantages and shortcomings: sophisticated capacities approaches, rudimentary capacities approaches, potentiality approaches, and relational approaches. One conclusion that can be drawn from the preceding discussion is that, as things stand, there is no philosophical consensus about the scope of our moral obligations. Ultimately, I shall argue that an advantage of an expressivist approach is that it can make good sense of the variability of intuitions that moral agents have concerning the properties or relations that confer moral status. In this section, however, I want to focus instead on a kind of uniformity underlying the views I have been examining. Despite the substantive differences between them, the theories of moral status examined so far share several commitments which hang together to form a philosophical picture. In this section I shall first make explicit these shared commitments. Then, in the following section, I shall introduce two accounts of moral status which reject them.

I shall use the term *moral individualism* as a label for the set of pre-theoretical commitments which underly standard or orthodox approaches to the grounds of moral status. This expression was employed by James Rachels, who uses it to denote "a thesis about the justification of judgements concerning how individuals may be treated. The basic idea is that how an individual may be treated is to be determined, not by considering his group memberships, but by considering his own particular characteristics" (Rachels 1990, 173). For Rachels, moral individualism's attraction lies in its ability to respect the species neutrality requirement, while promoting a principle of equality according to which if we treat A in a way that differs from how we treat B, then we must point to some (relevant) characteristics of A and B *qua* individuals.

As Susana Monsó and Herwig Grimm have pointed out, although Rachels' original formulation of moral individualism takes the form of "a metaethical principle governing how to justify differences in treatment across individuals" (Monsó and Grimm 2019, 1057), it is sometimes (erroneously) construed as "a doctrine about moral status" (1057).[44] While I agree that moral individualism is not itself a theory of moral status (I take it to be a set of pre-theoretical commitments on which various theories of moral status are premised), I contend that it involves more than just a metaethical principle. Unless more is said about what counts as an individual's "particular characteristics," it is hard to discern the position with which moral individualism is supposed to contrast. Rachels' formulation implies that "group memberships" are not to be considered as "belonging to an individual" in the relevant sense. However, it is not obvious that this suggestion is helpful or even warranted. Surely my being a member of the

[44] For example, Jeff McMahan (2005) defends moral individualism as an account of moral status, whereas Alice Crary (2010) attacks it so construed.

human species or belonging to the family that I do are just as much characteristics of who I am as the fact that I have brown eyes or can speak English. In order for moral individualism to avoid vacuity, it requires (or presupposes) a principled way of distinguishing those properties or characteristics which count as "belonging to an individual" and those which do not.

Perhaps unsurprisingly, many writers insist that it is only an entity's *intrinsic properties* (as opposed to its *extrinsic* or *relational properties*) which can properly confer moral status.[45] Given this distinction, it might seem that the proper contrast with moral individualism is something like moral relationalism—the view that how an individual ought to be treated is to be determined on the basis of the relations in which it stands. Todd May has recently defended a position along these lines, opting for an ecumenical view which recognizes that both intrinsic properties and relationships give rise to sets of mutually irreducible moral reasons (May 2014). Yet, as others have pointed out, "This distinction [between individualism and relationalism] appears to be misguided, for the simple fact that relations are also characteristics of individuals, and so it is unclear why moral relationalism should not be viewed as a form of [moral individualism]" (Monsó and Grimm 2019, 1057). Moreover, even if an intelligible and compelling distinction between intrinsic and extrinsic properties could be drawn, this alone would not tell us which intrinsic properties are relevant to moral decision making. A person's sentience and her mass are both intrinsic properties (if anything is), yet nobody would claim that the latter bears on how she should be treated morally. This suggests that the moral individualist

[45] An intrinsic property is one that an individual possesses in and of itself. Relational properties are ones that individuals hold in virtue of their relationships to other things. As Warren explains, "An entity's intrinsic properties are those which it has, and which it would be logically possible for it to have had even if it were the only thing in existence. By contrast, its relational properties are those which it has, but which it is not logically possible for it to have had were it the only thing in existence. Life, sentience, and the capacity for moral agency are in this sense intrinsic properties, whereas being a grandmother, or a recently naturalized citizen of Canada, are relational properties" (Warren, 122-123).

requires a way of distinguishing among those properties which "belong to an individual," that is those which are morally relevant from those which are not.

Another way of putting this point is that moral individualism requires a way of specifying those properties or relations which are capable of generating agent-neutral reasons for moral obligations. Often, I think, it is simply assumed that only intrinsic properties can do this.[46] Whether or not such an assumption is correct is not my main concern here. Rather, by drawing attention to this primary motivation for moral individualism, I hope to elucidate several other commitments which tend to accompany it.

Above, I claimed that moral individualism constitutes a philosophical picture. By this I mean that it involves a cluster of pre-theoretical commitments which help define the kinds of questions and problems that appear salient, and provide the criteria by which solutions are deemed adequate. While they need not logically entail one another, I believe that theories of moral status predicated on moral individualism (as I have just characterized it) will typically embrace a number of additional commitments, which I shall call: (i) attitude independence, and (ii) rationalism. Allow me to examine these commitments and explain how I think they are related.

> *Attitude Independence*: whether or not an entity has moral status is a matter of its possessing certain properties or standing in certain relations independently of the attitudes that human beings happen to hold towards that entity.

This is a commitment that can be associated with what Sharon Street has called *realist theories of value*, according to which, "there are at least some evaluative facts or truths that hold

[46] One reason for this could be that something's having *intrinsic value* is thought to supervene on its intrinsic properties (Delon 2016, 371). Thus, another powerful intuition behind moral individualism might be that an entity can only have moral status by virtue of those properties which make it intrinsically valuable.

independently of our evaluative attitudes" (Street 2006, 110). Most of the philosophers whom I have been discussing, for instance, take it for granted that an adequate account of the grounds of moral status will identify a set of properties or relations which confer moral status *independently* of the practices or evaluative stances that moral agents happen to take.[47] Given their shared commitment to moral individualism, this should not be surprising. If one believes that moral status must be grounded in properties which generate obligations for *any* and *all* similarly situated moral agents, it follows that these properties must generate those obligations independently of any particular agent's attitudes. This focus on agent-neutrality leads individualists to downplay (or ignore entirely) the significant of the social and cultural contexts in which moral status is conferred.[48] I take it that this commitment to attitude independence is closely related to a second commitment:

> *Rationalism***:** the prioritization of principles over sentiments in guiding how we should think about moral status.

If an entity's having moral status depends on attitude-independent facts about it, then it seems plausible that the philosopher's role is to determine what those facts are and to derive from them principles that can guide ethical decision-making. According to this view, failures to recognize which entities do in fact possess moral status are typically regarded as failures of rationality,

[47] Some of the philosophers discussed in 4.3 reject the attitude independence commitment while still subscribing to the core moral individualist idea that certain non-moral properties or relations provide *agent-neutral moral reasons*. For example, Korsgaard's constructivism is premised on the idea that there cannot be value independently of valuers (thus, it rejects attitude independence), but it still aims to construct an account of how there are features of A's *being a valuer* that provide *anyone* with reasons to treat A with certain forms of respect.

[48] There are, of course, exceptions to this tendency within moral individualism. For example, although Kittay accepts that biosocial relations can be the source of agent-neutral reasons, she does not think that these can be understood without careful attention to the social and political contexts in which they are situated (Kittay 2005).

rather than deficiencies of imagination, sentiment, or empathy.[49] Consequently, moral individualists believe that an adequate theory of moral status would serve as a theoretical corrective to moral practice by helping us discern to whom we *really* have moral obligations. Although it is seldom stated explicitly, this way of thinking presupposes that moral progress involves an increased adequation of our beliefs to some antecedently existing moral reality.

Although these commitments may manifest themselves in different ways, I believe that they are widely accepted by those theorists considered above. Despite disagreement about which properties or relations actually ground moral status, most of the writers engaged in these debates do not disagree about what those properties are supposed to do—namely generate agent-neutral obligations, which can be used to develop a set of principles to guide moral conduct. These principles, it is assumed, are rationally obligatory and would serve as a much-needed corrective to present practices. And while moral individualism represents the orthodoxy within the literature on moral status, it is not without its critics. Indeed, a growing number of *non-standard accounts of moral status* reject moral individualism—either in part or entirely. In the next section, I shall present an overview of these accounts. It will help to have a way of referring to the different kinds of disagreements about the grounds of moral status. In what follows, I shall describe disagreements about the grounds of moral status internal to moral individualism as *first-order* disagreements about moral status. First-order disagreements concern *which* properties or relations ground moral status claims. I shall refer to disagreements between moral individualists and their critics (discussed in the next section) as *second-order* disagreements. These latter

[49] It is a commonplace among the theories examined in the previous section to regard “received” or “traditional” moral outlooks as irrational or inconsistent. David DeGrazia, for instance, describes those who deny moral status to non-human animals as seeing things “through the distorting lens of prejudice” (DeGrazia 1996, 44).

philosophical disputes concern not so much the grounds of moral status but the nature, scope, and aims that a theory of moral status ought to have in the first place.

5.4.2 Critics of Moral Individualism: Pluralism and Humanism

Mary Anne Warren and Elizabeth Anderson have both advanced what might be described as *pluralist* or *contextualist* approaches to questions of moral status. Both philosophers deny that moral status can be grounded on the basis of a single criteria and insist that questions about the nature and scope of our moral obligations are highly contextual.

The central targets of Warren's book are uni-criterial accounts of moral status: approaches which identify a single intrinsic or relational property as the basis of moral status. Warren argues that while these forms of moral individualism capture some important features of our moral decision-making, none of them can be given ultimate priority or serve as both necessary and sufficient conditions for the grounds of moral status. Instead, Warren advances a pluralist, *multi-criterial* approach on which "[t]hree intrinsic properties—life, sentience, and moral agency—are directly relevant to moral status, each in a different way" (Warren 1997, 148). At the same time, our embeddedness within complex social and ecological systems and our relationships with others generate irreducible moral obligations. Given the wide range of things to whom moral agents may have moral obligations and given the variety of reasons for such obligations, Warren is skeptical that we will be able to come up with a "simple formula" for deciding who or what has moral status (173). Instead she puts forward several defeasible and holistically applicable principles of moral status intended to guide our deliberative practices.[50] In

[50] These principles include: the respect for life principle, the anti-cruelty principle, the agent's rights principle, the human rights principle, the ecological principle, the interspecific principle (i.e., "*non-human members of mixed social communities have a stronger moral status than could be based upon their intrinsic properties alone"* (Warren 1997, 168), as well as the transitivity of respect principle (i.e., when feasible, moral agents ought to respect each

doing so, Warren aims to do justice to the intuition that there is something distinctive about our moral obligations towards other human beings (perhaps owing to the centrality of interpersonal relationships to our moral self-understanding) and non-human animals, while allowing for the possibility that non-sentient entities such as rivers or ecosystems can have moral status.[51]

In a similar spirit, Elizabeth Anderson has argued for a kind of value pluralism when it comes to negotiating disagreements between proponents of animal welfare, animal rights, and environmentalism. Echoing John Dewey's "Three Independent Factors in Morals", and William James's "The Moral Philosopher and the Moral Life", Anderson contends that tensions between these competing perspectives arise because each begins by taking a different value (or set of values) as central.[52] For Anderson, "while each perspective has identified a genuine ground of value, none has successfully generated a valid principle of action that does justice to all the values at stake. The plurality of values must be acknowledged" (279).

One consequence which Anderson draws from this value pluralism is that the argument from species overlap (to which many advocates for animal rights appeal) relies on misleading

other's attributions of moral status). It might be thought that in advancing principles, Warren's position is more closely aligned with moral individualism than I am presenting it. One reason for reading her as opposing individualism, is (in addition to her commitment to contextualism) her sustained criticisms of views which neglect the importance of moral sentiments and emotions (74-6, see esp. chapter 5).

[51] According to Warren's "ecological principle", "living things that are not moral agents, but that are important to the ecosystems of which they are a part have [within the limits of Warren's other principles] a stronger moral status than could be based upon their intrinsic properties alone" (Warren 1997, 166). This allows for (though does not require) the possibility that we have moral obligations towards "water, air, plant and animal species, or other elements of the biosphere that are neither living organisms nor sentient beings." (167). That being said, Warren stops short of claiming that it is "mandatory to accord moral status to entities that are neither sentient nor alive. Because such entities cannot be harmed in the ways that living things and sentient beings can, it is implausible to insist that our obligations regarding them must be understood as obligations towards them" (167).

[52] In particular, Anderson suggests that "The animal welfare perspective originates in our sympathetic reactions to animals. The animal rights perspective originates in our respect for animals, our sense that their independent perspectives make claims on us that we ought to heed. It also, although it does not want to admit it, trades on our esteem for animals. The environmentalist perspective originates in our wonder at and awe of nature, conceived as an interconnected system of organisms, as well as in our admiration for individual animals" (293). All of these values are important, yet they are qualitatively different.

oversimplifications about the complexities of human and non-human lives and fails to consider the social and historical relationships within which it makes sense to attribute rights in the first place (Anderson, 2004, 280, 289). For example, rather than simply grant non-human animals rights on the basis of the supposedly morally relevant properties they possess, Anderson argues that granting rights and moral status to non-human animals is only tenable when their interests are compatible with our own, and when they are capable of reciprocal accommodation. This means that granting rights and moral status will be not only species-dependent, but also dependent on historically contingent facts about human beings. For this reason, Anderson concludes that,

> "there is no single criterion of moral considerability, and that what rights should be extended to a creature depend not only on its individual intrinsic capacities, but on its species nature, its natural and social relations to the moral agents to whom rights claims are addressed, and the social and historical background conditions applicable to the moral agents themselves" (290).

What sets Warren and Anderson's accounts apart from the standard approaches to moral status is their rejection of the attitude independence assumption (both resist the idea that attributions of moral status can be made without a view to the social conditions within which they occur). Although they both accept that moral principles play an important role in our lives, they insist that this role is always contextual, and reject the idea that a focus on moral principles should be allowed to eclipse the importance of moral emotions and sentiments.[53]

Another family of views—which I shall call *humanist approaches*—departs from the standard individualist approaches listed above. Like the pluralist accounts, these humanist views

[53] Of the views discussed in this chapter, Anderson's value pluralism has the most in common with the genealogical account of moral status that I develop in the next two chapters. Whereas Anderson places the emphasis on the need to account for a multitude of *values*, my account places the emphasis on the need to account for various *deliberative strategies* that we employ when arriving at moral status ascription. On my view, the need for this kind of pluralism stems from the fact that our moral practices are extremely complex and present us with diverse problems requiring a variety of modes of reflection and responsiveness.

are characterized by a kind of radical contextualism according to which the meaning of moral concepts and obligations are inextricably bound up with the practices, standpoints, and forms of life within which they operate. To simplify a bit, there are roughly two strands comprising this humanist position. One strand, exemplified by the work of Cora Diamond, Alice Crary, Stephen Mulhall, Timothy Chappell, and Raimond Gaita, takes its point of departure from Wittgenstein. The other strand takes its starting point from the work of Bernard Williams.[54] Common to both forms of humanism is the insistence that the concept of a "human being" is a thick ethical concept which plays an indispensable role within our moral lives.[55]

As I understand them, the Wittgensteinian humanists develop a general and a particular point, both intended to undermine some of the moral individualist's key assumptions. The general point takes aim at the attitude independence assumption by essentially turning moral individualism on its head. *Pace* moral individualism, we do not come to recognize who or what has moral status by simply noting whether it has the relevant non-moral properties (e.g., sentience, rationality, etc.), rather, our ability to recognize and attribute those properties itself depends on an established background of ethical dispositions and responses. Alice Crary makes this point by considering the process through which one learns how to attribute mental states to

[54] Williams' association with humanist accounts of moral status is mostly due to a posthumously published essay, entitled, "The Human Prejudice", in which he argues that the fact that someone is a human being can serve as a legitimate reason for our having obligations towards them. In particular, Williams contends that arguments against speciesism (and here is primary target is Peter Singer) presuppose an implausible ideal of impartiality which renders our projects and concerns unintelligible (Williams 2006). For a critique of Williams' view, see Singer (2009) and Savulescu (2008). For a defense and elaboration on his ideas, see Diamond (2018), and Grau (2016).

[55] This notion often appears as the main point of contention between moral individualists and humanists because it violates the species neutrality requirement to which individualists typically subscribe. For example, see Peter Singer's critique of Bernard Williams (Singer 2009, 572). Without denying the importance of this theoretical disagreement, I believe that the point is subsidiary to the humanists' more central objection to individualism as a set of pretheoretical commitments (which I have characterized above). This is because there is nothing inherent to moral individualism (as I have described it) which makes the species neutrality requirement rationally necessary. Rather, most individualist just happen to share the intuition that species membership is morally arbitrary. On my view, a moral individualist could very well accept that "being human" properly grounds moral status, so long as they were able to argue that having the property provided agent-neutral reasons supporting certain moral obligations.

other beings.[56] We learn what pain is, for example, not by looking inward and observing some private (and perhaps immaterial) sensation which we then project onto others; but, rather, we acquire the concept through others' responses to our behavior and by learning to appropriately respond to theirs (Crary 2010, 25). This process is successful insofar as it:

> imparts an appreciation of the (sometimes helpful and often horrible) role of pain in our lives and at the same time positions us to see the relevant patterns of behavior, in a manner that presupposes an appreciation of this role, as having kinds of importance in virtue of which they essentially invite certain responses. It is a learning process that equips us to think and talk about the lives we lead with pain using categories - such as, e.g., "groaning," "moaning," "grimacing," and "straining" - that are not only physically irreducible but also normatively nonneutral in the sense that the idea of the appropriateness of particular modes of response is internal to them (26).

But if this is right, then attributing pain to others involves more than simply recognizing some normatively neutral fact about them. Rather, doing so presupposes a practical grasp of the kinds of appropriate responsiveness and attention to pain-behaviour. Whereas the moral individualist regards the capacity for pain as providing agent-neutral reasons for certain kinds of normative responses, Crary can be read as insisting on the normative priority of our practices, such that our ability to understand what pain is, is itself "necessarily guided by a conception of the kinds of things that matter in lives like ours" (26).

Timothy Chappell advances a similar line of argument with respect to our attributions of personhood. In many cases, he contends, our taking certain creatures to be persons occurs prior

[56] While Crary's discussion is motivated by an account of mental categories which she gleans from Wittgenstein's later work, Andrew Gleeson makes a similar point about our use of thick ethical concepts more generally (Gleeson 2008). With respect to the concept of *murder* Gleeson writes that, "If we say its evil consists in the grief inflicted on survivors, then we will not distinguish it from death by natural causes (since it also inflicts grief) unless the grief is specified as being grief over the victim's being murdered. The murder only produces these distinctive responses (ones to murder) because the people concerned take murder to have that meaning (take it to be murder) independently of their responses. We travel a circle back to murder as an irreducible moral concept. The same goes for rape, bullying, or racism. These humiliate because of what, independently of natural effects (responses of hurt, fear, anger etc.), they mean. What is done to us is not humiliating because we feel humiliated; we feel humiliated because what is done to us is humiliating (Gleeson 2008, 160-161).

to our attributing to them any of the characteristics which many individualists typically associate with personhood. According to this *proleptic* view:

> we do not look for sentience or rationality or self-awareness in a creature as a test to decide whether or not that creature counts as a person. It is the other way round. Having once decided, on other grounds, that a creature is a person, we know that this makes it the kind of creature that is likely to display sentience, rationality, self-awareness, and the rest of the personal properties. Hence, we look for displays of these properties from the creature. That is to say, we treat it as a person in advance of any such displays (Chappell 2011, 7).

Paradigmatically, Chappell contends, this is evident in our attitudes towards infants and small children. A parent, after all, "does not start out by treating her child as an inanimate object, like a sofa or a refrigerator or a rubber plant, and grudgingly consent to adjust her attitude toward it, one little step at a time, only as and when it proves itself more than inanimate by passing a succession of behavioral tests" (7). On the contrary, her tendency to talk to her infant, and to treat him as though he were a creature who can "reason, respond, reflect, feel, laugh, [and] think about itself as a person" are evidence of her treating the child "proleptically, in the light of the ideal of personhood" (8-9).

If the Wittgensteinian humanist's general point is that our ability to apply certain concepts (such as mental qualities or personhood) presupposes a background of norm-governed attitudes and responses, their particular point consists in applying this reminder to the concept of a *human being*. On the picture recommended by moral individualism, the fact that something is a human being is (typically) *irrelevant* to the kinds of responses that moral agents owe it. What is relevant, as we have seen, is supposed to be its possession of other properties, such as the ability to suffer, or moral agency. In her 1978 paper "Eating Meat and Eating People", Cora Diamond suggests that this picture gets things exactly the wrong way around. She writes,

> We can most naturally speak of a kind of action as morally wrong when we have some firm grasp of what kind of beings are involved. But there are some actions, like giving people names, that are part of the way we come to understand and indicate our recognition of what kind it is with which we are concerned (Diamond 1978, 469).

Grasping the meaning of the concept *human being* involves, among other things, learning that they are the kinds of things that have names, whose achievements should be celebrated and whose losses can be lamented and mourned. In particular, we learn that human beings are not the kinds of things to be eaten—even after they have died or have had a limb amputated. Someone who failed to grasp these aspects of this thick ethical concept (and the deep reverence for human life they suggest) would be hard-pressed to explain the point of our holding birthday parties and funerals for human beings. Again, when it comes to understanding our moral obligations, this humanistic point can be read as an insistence on the normative priority of social practices. As Diamond explains, these and other activities,

> are all things that go to determine what sort of concept 'human being' is. Similarly with having duties to human beings. This is not a consequence of what human beings are, it is not justified by what human beings are: it is itself one of the things which go to build our notion of human beings (Diamond 1978, 470).

In other words: our moral obligations towards X are not grounded in the properties which X essentially has. Rather, our moral obligations are part of what constitutes our concept of what an X is. From this perspective, any attempt to set aside the category of "human being" will inevitably lead to an impoverished view of our moral practices given its inextricable connection to our understanding of moral obligations. More generally, Diamond thinks that the entire approach of grounding our moral obligations on a set of non-moral properties fails to make sense of how we should treat animals. This absence is glaring when we consider that an implication of moral individualism is that "there is absolutely nothing queer, nothing at all odd, in the vegetarian eating the cow that has obligingly been struck by lightning" (Diamond 1978, 468).

Despite being minority positions within the literature on moral status, there are clear advantages to both the pluralist accounts offered by Anderson and Warren as well as those advanced by the humanists. By playing up the contextual nature of ethical decision making and downplaying the stringency of moral principles, these views can avoid the unacceptable implications or "bullet-biting" that various forms of individualism tend to invite.[57] Moreover, by insisting on the practical priority of moral deliberation, these views are less likely to appear as invoking external constraints, which arguably fail to connect up with moral agents' motivations and practical identities (Gruen 2017). This point is connected to their emphasis on moral emotions and the importance of art and literature not only in shaping our moral identities but also holding the promise of moral transformation.[58]

That being said, these views face some serious challenges. From the standpoint of the more orthodox positions, it is hard to see how non-individualist frameworks even count as adequate answers to the key philosophical questions: namely how to determine the scope of our obligations. That is, these positions do not appear to offer clear determinations about who or what has moral status. Perhaps even more seriously, one might contend that in demoting

[57] In his review of Jeff McMahan's *The Ethics of Killing,* Stephen Mulhall writes, "Two-thirds of the way through this dense, involved and exhausting book, its author acknowledges that his views about the nature of persons have the following implication. Suppose that a woman, without family or friends, dies giving birth to a healthy infant. At the same hospital there are three five-year-old children who will die if they do not receive organ transplants, and the newborn has exactly the right tissue type. If Jeff McMahan's theory is right, it is morally permissible to 'sacrifice' the orphaned infant in order to save the other three children. We can hold off for a moment on the question of why his theory has this implication. The simple fact that it does, as he points out, appears to be a *reductio ad absurdum* of his position. And McMahan has the grace to confess that he 'cannot embrace' this implication 'without significant misgivings and considerable unease'. But he embraces it nevertheless, since he continues his examination of abortion and euthanasia for a further 150 pages, still drawing on the same account of the nature of persons that led to the apparent *reductio* in the first place - as if simply acknowledging its existence constituted a sufficient settling of accounts with it" (Mulhall 2002, 1).

[58] When it comes to the question of what is "involved in trying to show someone that he ought not to eat meat" once one has abandoned the moral individualist's rationalistic strategy, Diamond turns immediately to the power that poetry has to extend to animals those "modes of thinking characteristic of our responses to human being" (Diamond 1978, 472-4).

principles and playing up the significance that imagination and emotions in our practical decision making, these views run the risk of leading to serious inconsistencies and shirking of moral responsibility. At worst, they seem to leave us unable to make sense of reasoned disagreement.

Unfortunately, the dialogue between individualists and their critics has not always been productive. Proponents of the kinds of pluralist and humanist approaches discussed in this section often charge moral individualists with grave oversimplifications and distortions of moral problems.[59] Moral individualists often respond by charging their critics with attempting to substitute empty rhetoric for substantive and serious philosophical argumentation.[60] As a result, the disagreement between moral individualists and their critics begins to take on an air of intractability. Faced with this possibility, there are least three lines of response that one might undertake.

One might hold that one of the positions is simply mistaken, or perhaps that both are. This is how proponents of either side tend to present the issue. The problem with this angle is that it overlooks clear advantages to both ways of approaching questions of moral status and fails to explain why much of what both individualists and non-individualists offer seems plausible. Anyone who adopts this response to the problem of intractability would, therefore, need to offer a theory of error—an explanation of why the other side gets things wrong. Another possible response would be to argue that these divergent approaches to questions of moral status stem

[59] Diamond is troubled by the "obtuseness" and "shallowness" of Singer and Regan's discussions of vegetarianism (Diamond 1978, 468), which she finds to be inevitably "self-destructive" (471). Similarly, Crary argues that the moral individualists' defenses of animal rights or well-being are "grounded in a confused picture of moral relationships among human beings" (Crary 2010, 22).

[60] McMahan writes, that "what emerges from a careful exploration of the writings of these Wittgensteinians is that the notion of a human being, divorced from its basic biological meaning, has no determinate substance there at all: it is merely a rhetorical ornament that takes over the function of persuasion when the argument runs out" (McMahan 2005, 379).

from incompatible metaphilosophical intuitions or commitments. Perhaps, the substantive theoretical differences simply reflect a divergence of what William James called tough-minded and tender-minded philosophical sensibilities: one side welcomes the impersonal nature of moral obligations as a central feature of the moral landscape; the other does not. What more is there to say? While this approach has the advantage of at least providing an explanation for the disagreement between individualists and non-individualists, it still leaves much to be desired insofar as it leaves the philosophical debate at an impasse and seems to foreclose the possibility that progress could be made in our thinking about these issues. A third response to the problem of intractability is to advance a broader explanation, according to which both the individualist and non-individualist theories of moral status reflect a kind of functional diversity within our ethical practices. Rather than regarding their differences as the product of an intractable disagreement, the upshot of this approach would be to show that they serve different purposes.

A central motivation of the expressivist position which I advance is that it promises to offer this third sort of response to the problem of intractability. By this I mean that it can shed light not only on the *first-order* disagreements between the various orthodox individualist accounts, but also on what I am calling *second-order* disagreements about moral status—disagreements between moral individualists and humanists. In order to motivate this claim, allow me to offer a brief description of the view which I shall develop in the next chapter.

Putting to work the methodological lessons gleaned in Part One (i.e., adding to the global expressivist's methodological toolbox the insights of pragmatic genealogy) the expressivist account that I advance will begin with a state-of-nature genealogy involving an idealized model of our moral practices. On the basis of this model, I shall propose a set of hypotheses aiming to explain how conceptual practices involving moral status would have emerged. I contend that the notion of moral status performs two clusters of functions: first, attributions of, or claims about moral status discharge an important generalizing function within moral practice, serving as a convenient shorthand for our moral obligations *as a whole* (as well as the reasons for those obligations). This generalizing function, in turn, helps moral agents give voice to their practical or moral identities, and facilitates deliberation about which identities to adopt or relinquish. Second, I shall suggest that the concept of moral status has an important progress-articulating function, which enables us to describe how changes in our practices would constitute moral improvement.

In line with the pragmatic genealogist's methodological advice, the next stage in my proposal will involve de-idealizing this state-of-nature model, by suggesting ways in which conceptual practices involving moral status have been elaborated and diversified. For example, historical developments have not only given rise to a proliferation in kinds and number of practical identities that a person may adopt within her life, but there have arisen varied, and occasionally competing conceptions of moral progress. By tracking these developments, an expressivist approach can highlight the functional diversity that the notion of moral status has taken on. This, I claim, provides an explanatory pattern which can help overcome the problem of intractability. From the standpoint of our present moral practices and theoretical landscape, we encounter what appear to be incompatible theories about the grounds of moral status: not only are there *first-order* disagreements about which properties ground moral status; but we find *second-order* disagreements about the viability of moral individualism. One of the goals of expressivist framework is to make sense of this diversity—by showing how the notion of moral status has come to take on different functional roles in light of changing circumstances and in response to various practical pressures.

In their recent essay on Edward Craig's genealogy of the concept of knowledge, Martin Kusch and Robin McKenna describe this explanatory pattern in terms of "retrodiction." These authors suggest that a virtue of Craig's approach is that it "'retrodicts'—both central features of our knowledge-talk and many existing analyses of our concept of knowledge" (Kusch and McKenna 2020, 1058). In a similar vein, the expressivist account of moral status which I advance puts forth an explanatory hypothesis about the central functions that the concept plays (given generic human needs) which is, in turn, able to "retrodict" not only central features of our moral practices, but also the kinds of theoretical difference which appear to be intractable. A de-idealizing genealogy which reveals, for example, how the concept of moral status has come to link up with evolving conceptions of moral progress, or has come to serve various roles within changing forms of moral deliberation, promises to give a wider perspective on the disagreement between moral individualists and their critics. What initially appear as an intractable (and perhaps, inexplicable) theoretical divergence, is now explained in terms of an underlying functional plurality within our ethical practices.

5.5 The Eliminativist Challenge

So far, I have suggested that the *problem of intractability* represents an important motivation for a genealogical account of moral status. In this section, I consider a second challenge for theories of moral status. Several authors have argued that we should drop the concept from our moral theorizing. These eliminativists—as I shall call them—argue that the concept of moral status is either confused or useless. If cogent, this objection would present an obvious challenge to my proposal: if the concept of moral status should be set aside, then it would seem superfluous to offer an expressivist account of it. In this section, I shall briefly consider the elimanitivists' arguments. While I believe that, in their current form, they do pose an unresolved problem for extant theories of moral status, rather than ruling out the plausibility of an expressivist account of moral status, I contend that the eliminativist challenge actually provides a motivation for it. Given its concern with functional and etiological questions, an expressivist *pragmatic genealogy* of moral status provides a direct response to the eliminativist's worry that the concept is useless or confusing,

Several theorists have called into question the intelligibility or usefulness of the concept of moral status along related terms like moral standing, or moral considerability (Rachels 2005; Sachs 2011; Horta 2017). In a 2005 essay, James Rachels argued that a careful consideration of the connection between facts about individual entities and our reasons for treating them in certain ways reveals that the notion of moral status is, at best, a dispensable one. Rachels points out that,

> Facts about people often figure into the reasons why they may or may not be treated in this or that way. Adam may be ejected from the choir because he can't sing. Betty may be given Prozac because she is depressed. Charles may be congratulated because he has just gotten engaged. Doris may be promoted because she is a hard worker (Rachels 2005, 167).

Yet it is a mistake, he contends, to think that there is some set of features or properties that someone possess, which underwrites facts about how they ought to be treated *as such*. This, however, is precisely what theories of moral standing seem to be committed to. Rather, according to Rachels,

> moral standing is always moral standing with respect to some particular mode of treatment. A sentient being has moral standing with respect to not being tortured. A self-conscious being has moral standing with respect to not being humiliated. An autonomous being has moral standing with respect to not being coerced. And so on. If asked, toward whom is it appropriate to direct fundamental moral consideration? we could reply: It is appropriate to direct moral consideration toward any individual who has any of the indefinitely long list of characteristics that constitute morally good reasons why he or she should or should not be treated in any of the various ways in which individuals may be treated" (170).

But if one accepts this line of thinking, it is hard to see what theoretical work "moral standing" is even doing. Instead of a "theory of moral standing" Rachels thinks that to make progress in applied ethics, all we really need is a theory of reasons, which would include three components. First, it would involve claims about the moral permissibility of certain *actions* (i.e., how certain entities should, or should not be treated). Second, it would specify *reasons*, that is, considerations that counts in favor of doing or not doing the action, which connects the action to a benefit or harm for the individual. And finally, it would identify *facts* about certain entities which help explain why certain courses of action either would or would not cause them harm (Rachels 2005, 170).

More recently Benjamin Sachs and Oscar Horta have not only echoed Rachels' point that the concept of moral status adds nothing to debates about the ethics of marginal cases, but have gone so far as to suggest that the concept is causing confusion.

Sachs points to the fact that expressions of the form "to have moral status is to *x*" are ambiguous as to whether they identify moral status with certain moral properties, or whether moral status is itself something from which the possession of moral properties can be inferred (Sachs 2011, 89-90). On the former understanding, he thinks, it would be parsimonious to just drop claims about moral status altogether, and to simply refer to the moral properties that entities have (e.g., that it has certain rights). On the latter understanding, if having moral status is sufficient for possessing certain moral properties, then the concept seems inert. When asked, for example, *why* a certain class of entities has the moral property of being worthy of respect, it is hardly satisfying to be told that those entities have moral status. The explanation we want, Sachs thinks, is that those entities have the relevant non-moral properties (e.g., that they can suffer).[61]

These considerations suggest that a concept's popularity is not always evidence of its utility. Indeed, a surprisingly large number of writers simply take for granted that the notion of moral status has a clear usage and is of theoretical value. If cogent, the arguments just presented would suggest that much, if not all of the literature on moral status rests on a philosophical confusion. Although I believe that the arguments adduced in support of eliminativism have merit, they are far from decisive.

First, the eliminativist arguments examined in this section appear to take moral individualism for granted. What Rachels, Sachs, and Horta have effectively argued is that the concept of moral status should be abandoned *because* it plays no clear role in facilitating inferences from an individual's non-moral properties to its moral ones. But this assumes that the only conceivable role for the concept of moral status to perform within moral theory is to help moral agents determine a list of properties, the possession of which would generate agent-neutral reasons for action. Given this prior commitment to moral individualism, the eliminativists argue

[61] Similarly, Oscar Horta argues that "the idea of moral status does not shed light on the problem of how we should behave towards different individuals in different circumstances where the satisfaction of their interests is at stake" (Horta 2017, 900). Horta first considers cases in which attributions of moral status are taken to be consistent with the principle of equal consideration of interests. Theorists who want to uphold this principle, he thinks, can do so only by either going in for a kind of radical contextualism whereby "someone's moral status can change at different times, depending on how the weight of her interests varies in relation to the weight of the interests of others" (903), or by insisting that having moral status just means that one's interests should be taken into consideration. In the latter case, he thinks, learning that something has moral status does nothing to help us decide how its interested ought to be factored into consideration. Thus, one could say instead "that the reasons we have to behave towards different individuals depend on their features and the circumstances in which they are, without having to introduce some extra intermediate concept, be it status or considerability" (903). Alternatively, Horta considers theories of moral status which explicitly reject the principle of equal consideration of interests. One might, for instance, argue that the interests of certain beings—say, those with sophisticated cognitive capacities—ought to be given priority, or given greater weight when deliberating to do about situations when interests conflict. Horta rejects several attempts to support this claim, as either arbitrary or having unacceptable implications.

that one could simply talk instead of claims about non-moral properties and the obligations they do or do not generate.

As we have seen, however, there are theorists who reject moral individualism. This suggests that the concept of moral status may contribute something valuable to our moral discourse which the eliminativist's have overlooked. Mary Anne Warren, for instance, has argued that the concept can "be used to specify minimum standards of acceptable behavior towards entities of a given sort" (Warren 1997, 13), and to "establish moral ideals" which "create a conceptual space for supererogation, encouraging individuals to move beyond conformity to minimum standards of acceptable behavior" (14). Another possibility is that attributions of moral status are not simply interchangeable with claims that entities have certain properties, but indicate that agents have an obligation to engage empathetically with others, or that they have an obligation to factor those entities' interests into their deliberation.[62] In this respect, saying that some entity has moral status could be thought of as a success term—indicating our ability to take up its point of view—or, perhaps (as Warren's second point suggests) as a kind of exhortation to do so. Obviously these suggestions would require further elaboration, but their availability suggests that it would be too hasty to abandon the concept of moral status simply because of its apparent dispensability within the narrow deliberative space carved out for it within moral individualism.

Moreover, several philosophers who subscribe to moral individualism have specified a use for the concept of moral status which eliminativists appear to have overlooked. David

[62] Sachs briefly considers, but ultimately rejects a proposal along these lines, which he attributes to Lori Gruen (Sachs 2011, 94, note 18).

DeGrazia has anticipated the eliminativist's argument and offered the following reply. He writes that,

> One might claim that assertions of moral status are redundant, adding nothing to certain claims about our obligations and their grounds… Rather than asserting that cats have moral status, we might just assert, say, that we have an obligation not to harm cats needlessly, that cats can be harmed because they have interests (grounded in their experiential welfare), and that the obligation rests at least partly on their having such interests, which are thwarted when cats are harmed (DeGrazia 2008, 184).

DeGrazia concedes that in principle, "talk of moral status is redundant and can always in principle be replaced by other language" (184), but holds that it "furnishes a convenient shorthand for general assertions about our moral obligations to beings of different sorts and the grounds of those obligations" (184). In her recent book, Korsgaard reaches virtually the same conclusion, explaining that the concept of "moral standing"

> is a kind of stand-in, a kind of variable, for whatever it is that explains why we have obligations to the members of some group of entities, or more generally, for whatever it is that determines how we should treat the members of some group of entities" (Korsgaard 2018, 96).

What DeGrazia and Korsgaard are gesturing at, I take it, is that the concept of moral status plays a kind of generalizing function within our deliberative practices. *Pace* the eliminativists, the concept is not (necessarily) used to draw inferences about how to treat certain marginal cases, but rather, a way of making general assertions about the scope of our moral obligations and the reasons which underly them.

The eliminativist might respond that DeGrazia's and Korsgaard's suggestions are underdeveloped. If the issue has to do with the usefulness of moral status as a concept, then being told that it allows us to make generalizations or to act as a kind of moral variable, simply invites the further question about the point of those activities. Given that neither DeGrazia nor Korsgaard explains what work such generalizations is supposed to perform, one might think that the eliminativist's objection has just been set aside, only to re-emerge down the road. In the next chapter I shall offer a much more detailed elaboration of DeGrazia and Korsgaard's suggestion by exploring how the generalizing function of moral status talk allows moral agents to articulate and deliberate about their moral identities, and to voice conceptions of moral progress which would otherwise be difficult to put into words.

Far from providing a reason for dismissing an expressivist account of moral status, I submit that the eliminativist challenge represents both a constraint on and a motivation for an expressivist account. An expressivist approach needs to make sense of the role that the concept of moral status plays within our deliberative practices. But as it turns out, the kind of functional, subject naturalistic perspective I shall develop in the next chapter is well-suited to elaborate this kind of claim.

5.6 Towards an Expressivist Account of Moral Status: Its Challenges and Advantages

In this chapter, I suggested that recent debates concerning moral status can be divided into two categories. On the one hand, the more orthodox positions committed to what I called moral individualism, hold that moral status is grounded in an individual's properties—a position, I argued, that tends to involve further commitments to attitude independence and rationalism. On the other hand, non-standard views of moral status depart from some or all of those shared commitments. While these positions avoid some of the problems facing individualist accounts, I argued that they face challenges of their own. I then called attention to an unresolved problem within debates about moral status, *the problem of intractability*. This problem involves two dimensions. On the one hand, there are (seemingly) irreconcilable *first-order* disagreements about the properties or relations that ground moral status claims. On the other hand, there are *second-order* disagreement about how inquiry into moral status ought to proceed. To put it in Kuhnian terms, *moral individualism* and *moral humanism* represent two "incommensurable paradigms" concerning how philosophers should approach issues about the nature and limits of moral concern.

Next, I considered a second unresolved problem for theories of moral status, the *eliminativist challenge.* Eliminativist's claim that the concept of moral status offers no guidance for moral theory and practice. So far, I have suggested that there are reasons to think that an expressivist account of moral status can overcome these objections. However, the only way to show this conclusively is to construct such an account—a task I take up in the next chapter.

With this overview in mind, I want to conclude by explaining some of the motivations for and potential advantages of an expressivist position.

The first advantage has to do with metaphysical economy. As I noted in Chapter Two, traditional metaethical expressivists typically tout as a theoretical selling-point their ability to circumvent metaphysical questions about the nature of moral facts and their place within the natural world (Blackburn 1998). An expressivist account of moral status can claim a similar advantage. Although debates about the grounds of moral status are seldom (if ever) explicitly about a realm of "moral status facts," the naturalistic approach I am recommending arguably avoids metaphysical disputes that are caught up in inquiries into the grounds of moral status. As we have seen, many of the standard accounts outlined in this chapter require that one take a stand on certain metaphysical debates about the nature of personal identity over time, the nature of potentiality, consciousness, or the distinction between intrinsic and relational properties.

The widespread sentiment that questions about the grounds of moral status *necessarily* require metaphysical heavy lifting is well expressed by Jeff McMahan. In an in an interview about the topic of moral status, McMahan is asked the following questions:

> So, what you're saying is that before you can make a judgement about moral status, you have to understand the metaphysics of what it is to be a person. And a consequence of that is that most people aren't actually equipped to make judgements about moral status.

To which he replies:

> Unfortunately, I think that that's correct. These are issues about human beings (and other animals) whose nature is in some sense non-standard: embryos, foetuses, newborn infants, adults with certain cognitive impairments or radical deficits. These are individuals about whose moral status we should not have confident intuitions and confident moral views. Questions about abortion, the termination of life support, euthanasia, and so on, are really very difficult. We are right to be puzzled about these issues, and people who think that they know the answers and have very strong views about these matters without having addressed the difficult issues in metaphysics and moral theory are, I think, making a mistake (McMahan 2012).

While I am not claiming that these metaphysical debates are pointless or somehow mistaken, I am more diffident than McMahan that they are the sort of debates about which one might expect something resembling a consensus anytime soon. The advantage I am claiming for the expressivist approach is that it allows one to approach philosophical questions regarding moral status from a different angle, one which does not require one to take a position on these substantive metaphysical debates.[63]

In this chapter, I have suggested that philosophical debates about the grounds of moral status have an air of intractability to them. A second advantage of an expressivist account of moral status is that it allows us to approach both *first-order* and *second-order* disagreements about moral status from a novel perspective and in a way that promises to make sense of why

[63] As I mentioned in Chapter Two, this is not to say that the expressivist approach I am advocating is utterly devoid of ontological commitments. It is, after all, committed to the existence of human beings with certain capacities and needs, inhabiting familiar natural and social environments.

those disagreements take the shape they do. In section 5.4.2, I claimed that (when supplemented by the de-idealizing resources of philosophical genealogy), an expressivist account of moral status is able to “retrodict” central features of moral theory and practice, by attending to the functional diversity that the concept has taken on.

A third and final motivation for an expressivist account of moral status is to provide a response to the two challenges raised in this chapter. I believe that the kind of expressivism derived from an amended version of Price’s subject naturalism is particularly well-suited to meet both of them. On the one hand, beginning with a functional account of the concept of moral status promises to offer a direct response to the eliminativist’s claim that the concept of moral status does no work for us. On the other hand, the de-idealizing dimension of a subject naturalistic genealogy can reveal that this functional role is connected with other concepts and practices—in particular, with deliberating about our practical identities and articulating conceptions of moral progress. These connections can be elaborated in different ways, ultimately giving rise to competing pictures of moral deliberation. It is within this de-idealizing story that we may be able to locate a response to the realist challenge of making sense of disagreement.

CHAPTER 6

AN EXPRESSIVIST ACCOUNT OF MORAL STATUS

> If concept formation can be explained by facts of nature, shouldn't we be interested, not in grammar, but rather in what is its basis in nature? —We are, indeed, also interested in the correspondence between concepts and very general facts of nature. (Such facts as mostly do not strike us because of their generality.) But our interest is not thereby thrown back on to these possible causes of concept formation; we are not doing natural science; nor yet natural history—since we can also invent fictitious history for our purposes (Wittgenstein 1953/2009, PPF, §365).

6.1 Introduction

In the previous chapter, I identified two outstanding problems for those who employ the concept of moral status. The first, which I have called *the problem of intractability*, stems from the rift between moral individualists and their humanist opponents, and the fact that both approaches make intuitively compelling claims. The individualist appears to be on solid ground in asserting that an entity's capacities or relations are relevant to its having moral status. Similarly, there is much to be said for the humanist's proleptic account of personhood and their foregrounding of emotions and imaginative capacities as central to our moral practices. And yet we lack a satisfying explanation not only of why these two positions seem to diverge so sharply, but of a satisfying means of adjudicating between them. The second problem, which I have called *the problem of eliminativism,* is that the concept of moral status has no clear function and should be set aside. So long as one allows for talk of non-moral and moral properties—the argument goes—there is simply no need for an intermediary concept like moral status to facilitate inferences from the former to the latter.

In this chapter, I develop an expressivist account of moral status and argue that it can resolve these two problems. Rather than specify at the outset the properties or relations which ground moral status, the guiding insight of this approach is to first inquire into the concept's function and etiology.[1] There are two virtues of this approach. First it explains why moral individualists and humanists disagree by appealing to the fact that the concept of moral status is more functionally diverse than is typically recognized. Attending to the various roles the concept plays within moral practice promises to deliver a picture of moral status which leaves room for both individualism and humanism—albeit in relatively attenuated forms. On this view, where both positions go astray is in committing what Dewey called the philosopher's fallacy: a tendency to neglect the fact that all inquiry is selective and thereby to overgeneralize some theoretical abstraction (Dewey 1925/2010, 29). My aim is to suggest how both positions get part of the story right. Instead of viewing their disagreement as a product of some grave error, or as the product of incommensurable metaphilosophical commitments, it results from a tendency to focus on different aspects of our moral lives. To establish this claim, I need to show two things:

> *Functional plurality*: the concept of moral status plays multiple roles within our moral practices.

[1] A functional explanation of some practice X (or of some component of a practice) accounts for X's existence by looking to its characteristic roles, uses, or purposes, rather than trying to determine X's nature, essence, or meaning. In other words, functional accounts raise the question 'what does X do?' rather than the question 'what is X?' Functional explanations are ubiquitous in biology wherein features of an organism's morphology or behavior are explained in terms of the fact that they promote its survival or reproductive fitness. Similarly, functionalist approaches in the social sciences explain an institution's existence on the basis of the function that it serves. While pragmatic genealogy has affinities with these programs, I take its scope and explanatory ambitions to be much more limited. Unlike functionalist social science, for instance, I do not take the approach offerd in this chapter to be claiming that all social practices can or should be given a functionalist treatment. Rather, I see functional explanation as an underappreciated tool for approaching philosophical problems. Questions about function are distinct from questions of etiology, although the two can operate together. Etiology comes from the ancient Greek word αἰτιολογία, which means "an account of causes." Broadly speaking, an etiology is an account of how something came about. These types of explanations are familiar in a number of other contexts. In medicine, an etiology of a disease, is an investigation into its causes or origins. Etiologies also play an important role in mythology. The Book of Genesis, for instance, is rife with origin stories about various natural phenomena (e.g., rainbows came into existence as a sign of god's covenant with humans). Functional explanations and etiologies converge when an investigation into X's causes or origins makes reference to its functionality.

> *Theoretical correspondence*: these differences in function explain the disagreement between individualists and humanists (i.e., each theoretical framework takes as paradigmatic a different set of functional roles that the concept of moral status plays).

Second, an expressivist approach offers a direct response to the eliminativist. Drawing attention to the concept's overlooked functional diversity goes some way to responding to the claim that it is inert or redundant. In particular, I shall argue that certain forms of moral deliberation would be impossible without something like a concept of moral status. However, for reasons which I hope will become clear over the course of this chapter, the framework which I develop does not necessarily provide a full-blown vindication of our conceptual practices involving moral status.[2]

My methodological strategy draws from the amended version of Huw Price's subject naturalism which I developed in Chapter Four. This, recall, involved adding pragmatic genealogy to the subject naturalist's methodological toolkit. Before proceeding, I would like to ward off some potential confusions regarding my use of the term "expressivism." This label is meant to echo Price's global expressivism, which, recall, departs from traditional expressivism in its repudiation of the bifurcation thesis—the idea that one can draw a principled distinction between representational and non-representational domains of language (Price 2013, 30). Given that this chapter focuses so much on a genealogy of moral status, some readers might get the sense that the very idea of "expressivism" drops out of the picture, and that it would be more appropriate to call this proposal a "pragmatic genealogy."[3] One way for me to assuage this

[2] In part, this limitation stems from the fact that I think the concept of moral status is more functionally diverse than most philosophers have acknowledged. While I aim to show that any practices which we would be willing to regard as a distinctively "moral" would need *some conception* of moral status, I do not take my argument to vindicate any conception in particular.

[3] Another reason for this is that a considerable amount of contemporary expressivism is devoted to solving technical problems in formal semantics. While I do not engage with these types of issues here, nothing that I shall say rules out the possibility of developing a semantics for moral status claims.

potential worry is to say how the three theoretical positions identified in earlier chapters fit together in my account. First, at the most general level, my view begins by accepting—more or less—the tenets of Price's subject naturalism: roughly, that instead of making naturalistic sense of some target philosophical vocabulary by asking about the metaphysical status of its referents or truth-makers, one should look to the language-users who employ that vocabulary (Price 2013). Second, at the methodological level, this account could be described as a pragmatic genealogy of moral status. That is, it offers a two-fold framework—comprised of both idealization and de-idealization—which attempts to make sense of how human beings came to employ the concept and its role in our practices.[4] Finally, I describe the result of this inquiry as an expressivist account of moral status. This is because it looks to some area of discourse and finds that it is not straightforwardly fact-stating or descriptive—more technically, it makes no use of robust semantic properties or relations in order to explain how that vocabulary is used.[5]

So much for monikers. In section 6.2, I construct an idealizing state-of-nature model aiming to explain—functionally and etiologically—why our conceptual practices involving moral status would have emerged in light of generic human needs and purposes. As the kinds of creatures who are responsive to moral considerations, human beings find themselves in a basic moral predicament in which they are led to reflect upon, disagree and deliberate about the scope of entities to whom those attitudes are directed. The concept of moral status (or what is better

[4] In *Truth and Truthfulness*, Williams distinguishes between "fictional" or "imaginary" genealogy and genealogy that incorporates real, empirical history (Williams 2002, 39-40). I prefer the terminology of "idealizing" and "de-idealizing" given that it avoids the potentially misleading idea that a fictional (i.e., idealizing) component of a genealogy is somehow unreal or constructed arbitrarily. Following Kusch and McKenna (2020) and Queloz (2018), I think that an idealizing genealogy is best seen as an instance of the kind of exercise in explanatory model-building that is ubiquitous across the natural and human sciences.

[5] In Section 6.2.3 I explain how my account bears important similarities to the kind of logical expressivism advanced by Robert Brandom.

thought of as a proto-concept of moral status) allows for moral creatures to identify potential loci of moral and concern and to make generalizations about their obligations.

In sections 6.3, I proceed to the second stage of a pragmatic genealogy of moral status: *de-idealization*. My focus will be on two ways in which the *proto-concept* of moral status has come to be elaborated. First, by looking to the ways in which it allows for agents to deliberate about their moral identities. Second, by showing how the concept can come to play a role in articulating conceptions of moral progress. While I do not claim that these functions are exhaustive, they represent two directions in which the concept of moral status has been elaborated and transformed in light of changing practical exigencies and social conditions. Both processes of elaboration, I shall argue, illustrate how the concept of moral status begins to exhibit *functional diversity.* And by tracking this diversity, moreover, one can appreciate the diverging insights of both moral individualists and humanists. In other words, understanding how the concept of moral status comes to be used in different ways—for example, in facilitating different forms of reflection about moral identity, or in enabling distinct notions of moral progress to take shape—yields an explanation of the theoretical divergences depicted in the previous chapter, and thus, a solution to the problem of intractability.

Finally, in Section 6.4 I conclude by considering the eliminativist challenge in light of the preliminary framework set forth in this chapter. Although the pragmatic genealogy outlined here does much to defuse the claim that talk of moral status is redundant or useless, it does not provide a full-fledged vindication of our conceptual practices involving it. Such a vindication would require a more detailed historical elaboration which inquired into the emergence of our current practices in their historical specificity. In particular, it would need to investigate how the concept of moral status has become embedded within these practices and especially how it connects up with a more localized constellation of values, reasons, and meanings. In the following chapter I offer a sketch of what such an historical de-idealization might involve. I shall do so by considering recent debates about the legal and moral status of machines displaying intelligence.

6.2 Pragmatic Genealogy and the Moral State-of-Nature

In this section, I construct a state-of-nature model intended to explain why the concept of moral status would have emerged given a set of generic features of our moral practices. First, taking Edward Craig's and Bernard Williams's accounts as paradigms, I describe what a state-of-nature model involves and say why it is particularly well-suited to shed light on the concept of moral status. Second, I look to William James's "The Moral Philosopher and the Moral Life" as the basis for such a model of our distinctively moral practices. Third, against this Jamesian backdrop, I advance my central hypothesis: that the concept of moral status allows agents to engage in certain kinds of deliberative practices involving what I call *moral interpolation*. In order to extend conceptual practices to novel situations, to make judgments about new cases, and to resolve disagreements about normative indeterminacy, moral agents require a set of practical abilities, namely the ability to identify appropriate loci of practical and moral concern, and the ability to make generalizations about their judgments and obligations. As practices involving moral interpolation develop in complexity and sophistication, the need for a vocabulary which expresses these practical abilities becomes increasingly pressing. It is in light of these needs that the proto-concept of moral status could be expected to arise.

6.2.1 State-of-Nature Genealogies

State-of-nature explanations occupy a familiar place within political philosophy, especially in the social contract tradition running from Hobbes, Locke, and Rousseau.[6] Typically, a "state-of-nature" denotes a form of human life antecedent to social and political organization. The point of appealing to these—often entirely hypothetical—situations is twofold. Not only do they help make sense of why certain forms of social organization would have emerged, but they often perform a legitimating or justificatory role as well. The *locus classicus* of an explanatory, legitimating state-of-nature account can be found in Hobbes' *Leviathan*. Given certain environmental and social conditions (i.e., approximate physical and intellectual equality amongst people, scarcity of recourses) and a set of human dispositions (e.g., competition, diffidence, the desire for glory), Hobbes aimed to explain why it would be in their rational self-interest for people to give up their natural liberties and submit to the authority of a sovereign power (Hobbes 1996/1651). More recently, state-of-nature explanations have acquired popularity outside of political philosophy, where they have been used to throw light on concepts such as knowledge (Craig 1990), truthfulness (Williams 2002), and testimony (Fricker 2007). This repurposing of state-of-nature explanations has been accompanied by a methodological sophistication and self-consciousness. The present account follows suit by applying state-of-nature explanation to concepts central to our moral practices.

In Chapter Four, I suggested, following Martin Kusch, that the idealizing aspect of a state-of-nature model constitutes the first of two "stages" of a pragmatic genealogy—the second involving a process of de-idealization (Kusch 2009; Kusch and McKenna 2020; Queloz 2018a;

[6] See for example, Hobbes (1996/1651, chapter 13), Locke (1988/1689, the *Second Treatise* chapters 2 and 3), Rousseau (1987/1782).

2019). State-of-nature models *idealize* in at least two respects. First, in their depiction of human beings as possessing highly generic capacities and needs. And second, with respect to the social and environmental conditions in which those human beings find themselves. For both Williams and Craig, such a model involves social creatures who require information about a world which they experience from different perspectives. Abstraction and idealization are necessary in these contexts, in order to identify features of human psychology and sociality which could be given a claim to universality, if any of them can. Taken together, these features allow the genealogist to specify practical pressures on the basis of which some target phenomenon—that is, some feature of our conceptual practices—could be expected to arise. I shall use the term *practical eventuality* to denote this explanatory goal. That is, the pragmatic genealogist aims to show that some target practice is a practical eventuality, given a set of practical pressures foreseeable within the state-of-nature model. For example, Craig is ultimately concerned with explaining knowledge attribution in light of the basic need to flag good informants (Craig 1990, 11). Williams and Price, as I discussed in Chapter Four, both appeal to the generic social need to pool cognitive resources in order to explain how the normative character of truth and truthfulness could have come about (Williams 2002; Price 1988; 2003). All three writers, notably, posit something like an *epistemic* or *cognitive* state-of-nature—a set of basic conditions that humans face *qua* inquirers seeking to form beliefs.

For present purposes, several features of state-of-nature models (and pragmatic genealogies, more generally) are worth emphasizing. First, they are explicitly naturalistic. Williams puts this point in terms of the fact that a state-of-nature genealogy explains some complex (and perhaps, mysterious) phenomenon in light of simpler or uncontroversial features of our psychology (Williams 2002, 24-25).[7] As I argued in Chapter Four, it is precisely because Williams conceives of naturalism as an explanatory project directed at human subjects (as opposed to a reductionist one, grounded in natural objects) that his project can be read as complimentary to Price's. Recall, moreover, that for Williams, a state-of-nature is not a description of some specific historical period (e.g., the Pleistocene), but rather a kind of explanatory model (Kusch 2009, Queloz 2018a; 2019).[8] Though fictional, this naturalistic component places important constraints on state-of-nature model-building. A model's explanatory power depends on the fact that the conditions which it posits can be regarded as actually generic. As Craig puts it, state-of-nature explanations "work by identifying certain needs" and then showing that certain practices "are a necessary (or at least a highly appropriate) response to them" (Craig 1990, 89). Moreover, these kinds of explanations

> Will therefore be at their strongest when the human needs from which they start are the most practical, hence the most undeniable ones. This sets limits to what a 'state of nature' explanation can be good for. The less visible the practical significance, for us, of forming certain conceptions or operating a certain linguistic usage, the weaker the explanation (Craig 1990, 89).

A state-of-nature account which, for example, depicted culturally specific needs would have very limited explanatory power. Similarly, a state-of-nature model which was given full creative license in unpacking the etiology of some target social practice might be interesting, but it would

[7] Recall that Williams finds Hume's account of justice to be exemplary in this regard. For Hume, the reasons "involved in the virtue of justice" which are essentially "collective reasons for action", are derived from more primitive or basic "reasons of individual interest and limited sympathy" (Williams 2002, 33).

[8] Martin Kusch and Robin McKenna note an important connection between Craig's state-of-nature genealogies and model-building in the sciences. They write that "Craig's project has affinities with natural science in its method of hypothesis testing, the search for explanation, and a focus on evolution. Going beyond his wording, we would add model-building to the list: the building of simplified (and possibly even distorting) models of complex target systems" (Kusch and McKenna 2020, 1058).

be explanatorily useless from the perspective of naturalism since it would fail to pick out a plausible set of human needs on the basis of which that practice could be seen as a response.[9]

A second feature worth noting is that state-of-nature models yield functional explanations, which can be represented schematically as follows:

Target phenomenon P performs function F given circumstances C.

For example, Price explains the emergence of the truth norm (target phenomenon) by pointing to the fact that it ultimately improves collective epistemic resources (function), within a perspectival world in which humans require information (circumstances)(Price 1988; 2004). However, I noted in Chapter Four that a pragmatic genealogy's functionality generates one of its chief limitations, namely, that certain phenomena—especially those displaying what Queloz calls "self-effacing functionality"—are not reflectively stable when understood in purely functional terms (Queloz 2018b, 1-2). This is why pragmatic genealogies typically require a de-idealizing story which explains how some feature of a conceptual practices could be valued intrinsically.[10]

[9] The book of Genesis, for example, is full of etiological explanations, not just of natural phenomena (e.g., rainbows, the existence of salt deposits in the Levant) but of features of our psychology (e.g., the natural animosity which we humans feel towards snakes). In an obvious sense, these explanations fail to be naturalistic insofar as they depend on supernatural causes.

[10] Williams maintains that, in order to be vindicatory, a genealogy must be able to treat its target phenomenon as though it were of intrinsic or non-instrumental value, where this means that the target's value must make sense "from the inside" (Williams 2002, 91). In Chapter Four, I suggested that a problem for Price's account of truth was that it only portrayed the truth norm as an instrumental good. This, I argued, prompts worries about whether his account can be stable under reflection. The de-idealizing component of Williams' genealogy, by contrast, purports to show how truthfulness can be regarded as an intrinsic good, by situating within a constellation of other values, reasons, and goods with which people are likely to identify. For a critical discussion of Williams' conception of intrinsic value see Rorty (2002), Allen (2003, 364-367), and Koopman (2013, 68-73). For a defense of Williams' use of this idea see Queloz (2018, 9-14).

That is to say, a story which can connect that practice to other non-instrumental values or reasons that people are likely to hold.[11]

A further limitation, however, is that state-of-nature models tend to leave little room for functional diversity with respect to those conceptual practices they investigate. While nothing prevents the pragmatic genealogist from identifying within her state-of-nature model multiple functions that some target phenomenon discharges, for the most part, such variability will be limited and the exception to the rule.[12] This is because components of a conceptual practice will tend to take on new roles or functions as that practice gains in complexity and historical specificity—features which state-of-nature models aim to reduce through abstraction. Attending to functional diversity is, of course, a central goal of pragmatic genealogies, but it is one that they tend to achieve through de-idealization. As I shall discuss in Section 6.3, the genealogist's de-idealizing component identifies novel, historically contingent practical exigencies which will lead to modifications and extensions of a practice. It is within these dynamic "layers" that one is most likely to find functional diversity.[13] In *Knowledge and the State of Nature*, for example,

[11] Williams, for example, recognizes that the values surrounding the disposition of sincerity have "varied in different historical circumstances" (2002, 95) such that the kinds of reasons that we might give for regarding sincerity as a non-instrumental good are not the same as those which someone might have given two-hundred years ago. Nonetheless, these historically diverse conceptions—however different from each other—all regard sincerity "as having a value that goes beyond anything ascribed to it in the basic State of Nature story, where it first emerges as, roughly, the solution to a co-ordination problem" (95).

[12] Martin Kusch and Robin McKenna suggest that there is nothing about Edward Craig's conception of genealogy which entails that some target phenomenon (e.g., the concept of knowledge) must serve a single function within the state-of-nature model (Kusch and McKenna 2020, 1062). They go on to offer some possible "Craigian" responses to the objection that since it is possible to imagine multiple alternative functions that the concept of knowledge might have performed in the state-of-nature, that Craig's own account is rendered arbitrary (or defective in some other sense). First, they point out that any competing proposal will need to be judged on the basis of a number of theoretical virtues (e.g., exactness, fruitfulness, simplicity)—and that given the detail of Craig's account, this would require a lot of work. Second, even if turned out that the proto-concept of knowledge did have multiple (designated) functions, it could very well turn out that they are entirely complimentary, and do not stand in some kind of hierarchy (1063).

[13] For discussion of the idea that conceptual genealogies reveal different layers, see Dutilh Novaes (2015, 80).

Craig's initial hypothesis about the value of the concept of knowledge (i.e., that it allows inquirers to identify good informants) lacks the functional diversity associated with our present concept of *knowledge*. In its most primitive form (i.e., taken as what Kusch calls a "proto-concept") Craig suggests that the concept of knowledge would take on a subjective character. He writes, that in these conditions, "I am seeking information as to whether or not *p*, and hence want an informant who is satisfactory for my purposes, here and now, with my present beliefs and capacities for receiving information" (Craig 1990, 85). These subjective features limit the concept's application. It must allow speakers to pick out those who are accessible to them "here and now", there must be something about the informant that allows the speaker to recognize them as someone who is likely going to be correct about whether *p*, and "likely right" relative to the speaker's present needs and concerns as an inquirer.

These subjective dimensions, as Craig is wont to point out, prevent the proto-concept from being used in many of the ways that *our* concept of knowledge is. Kusch nicely summarizes these points as follows:

> [P]rotoknowledge differs from knowledge in that: (a) only the former is closely tied to testimony; (b) protoknowledge is not a fully public concept insofar as it is indexed to the capacities and needs of specific inquirers (1990: 90); (c) protoknowledge can be ascribed only to others but not to oneself; and (d) protoknowledge is not undermined by accident or luck: users of protoknowledge lack the intellectual sophistication to distinguish between accidental and non-accidental fulfilment of the conditions of protoknowledge (Kusch 2013, 66).

In other words, proto-knowledge lacks much of the functional diversity that the concept of knowledge enjoys.[14]

[14] To reiterate, this is not to say that a pragmatic genealogy is ill-suited to capture functional diversity, just that the idealizing component typically works against this. Craig introduces a process of de-idealization called "objectivisation" to explain how the proto-concept would have developed to resemble our own notion of knowledge (Craig 1990, 84).

A third feature of pragmatic genealogies is their sensitivity to the limitations of purely conceptual analysis. This point is especially worth noting because it helps explain why a pragmatic genealogy is an apt approach for shedding light on the concept of moral status. It is hard to overstate the centrality of conceptual analysis for 20th century Anglo-American philosophy. The idea that the best way of solving philosophical problems is to state necessary and sufficient conditions of concepts was, and still is, the *modus operendi* of many influential research programs, despite longstanding objections dogging such an approach.[15] Pragmatic genealogy is often framed explicitly as an alternative to pure conceptual analysis (Queloz 2019). For example, Craig, positions his own project of "conceptual synthesis" as an alternative to the so-called "S knows that P" analyses so familiar to twentieth century analytic epistemology.[16] Similarly, Williams and Price are both convinced that—aside from the banal equivalence scheme—philosophers would be better off inquiring into the function of truth rather than attempting to define or analyse it (Williams 2002, 63; Price 1988).

Although it is not always stated explicitly, many advocates of the accounts of moral status discussed in the previous chapter are hesitant to state the necessary and sufficient conditions of something's having moral status. Most moral individualists are sensitive to the fact

[15] For discussion of the limitations resulting from the so-called "paradox of philosophical analysis," see Blackburn (1984, 155). Miranda Fricker suggests that "analysis—understood as the attempt to achieve necessary and sufficient conditions—is not an appropriate method for any subject matters which have philosophically important features that are not necessary conditions. Such features will not figure in any strict definition, for the requisite trial by counter-example must ultimately eliminate them. And yet if these are explanatorily basic features, they are just the sort of thing that needs to be preserved in a philosophical account that aims to explain the nature of the practice in all its internal diversity. Successful analysis delivers the highest-common-denominator set of features of X; but where X is an internally diverse practice there is a significant risk that the highest common denominator will turn out to be very low, delivering an extremely thin account. In particular, it will not be capable of illuminating how the different forms of the practice are explanatorily related to one another" (Fricker 2016, 166).

[16] "Conceptual Synthesis" is a term found in the subtitle of Craig's book. In a nod to Carnap, he also refers to his project as a "practical explication of knowledge" (8), linking it to a form of naturalism which "spreads itself altogether wider than conceptual analysis" (8-9). This understanding of naturalism sees "man, his behaviour and institutions, as natural facts to be understood as the (broadly speaking causal) outcome of other natural facts" (9).

that they are only specifying sufficient *or* necessary conditions for the grounds of moral status.[17] Humanists, similarly, are not in the business of advancing analyses. Pragmatic genealogy represents an alternative to philosophical analysis which many of these authors would, therefore, have good reason to accept. But it is an alternative which can capture some of the insights so often foregrounded in philosophical analyses without giving up on the task of sense-making. In the previous chapter I noted that this is especially evident in the "retrodictive" aspect of Craig's work. Rather than view the repeated failures of philosophers to provide an analysis of knowledge as a failure *tout court*, he saw the various proposals as "data" in need of explanation. Just as Craig hoped to preserve the insight inherent in, for instance, truth-tracking, reliabilist, or causal analyses of *knowledge*, so too can a pragmatic genealogy of *moral status* preserve not only the insight that sophisticated capacities, sentience, relations, and potentiality, are intuitively relevant when it comes to questions about the scope of our moral concern (while allowing that these proposals admit of exceptions), but that humanists are on to something too. Not only are pragmatic genealogies, therefore, appealing tools for the ecumenically-minded, but they are especially well-suited to resolving the types of philosophical impasses of which the intractability problem is a token.

[17] David DeGrazia exemplifies what I take to be the cautious approach in refraining from proposing propose necessary and sufficient conditions for the grounds of moral status. As he puts it, "*To say that X has moral status is to say that (I) moral agents have obligations regarding X, (2) X has interests, and (3) the obligations are based (at least partly) on X's interests*" (DeGrazia 2008, 183). The third condition, as he points out, "is motivated by the possibility that some factor in addition to X's interests, such as the state of a moral agent's character, might also ground the relevant obligations" (184). In a similar vein, Shelly Kagan writes that, "According to personism your interests count more if you are a nonperson member of a person species or if you are a person — that is, regardless of whether your species is a person species. Since the intelligent dog is a person, its interests do count more. That is to say, either of two conditions suffices to have your interests count more (though one of these conditions may result in interests counting even more than with the other). Neither is necessary" (Kagan 2016, 13). S. Matthew Liao also expresses reluctance to offer an analysis of moral status, suggesting that his genetic-based approach is only meant to provide sufficient conditions (Liao 2010, 164). One exception is Elizabeth Harman, whose view is that "[a] thing has moral status just in case harms to it matter morally" (Harman 2003, 174).

6.2.2 The "Moral Life" as Moral State-of-Nature: James as Pragmatic Genealogist

For Craig and Williams, the state-of-nature model characterized something like a basic epistemic predicament. Since the present account is concerned not with explaining epistemic practices, concepts, values, or dispositions, but moral ones, it requires something slightly different. What is needed is a model incorporating the conditions endemic to a basic *moral* predicament. These generic features, in turn, can be used to identify those practical pressures against which the emergence of *something* like a concept of moral status makes sense. That is, through which it becomes a practical eventuality. I believe that one can find such a model in William James' essay, "The Moral Philosopher and the Moral Life".[18]

That this piece involves James' most explicit and direct foray into moral philosophy is largely undisputed. What it says about James' actual commitments as a moral philosopher, is, however, a matter of ongoing debate. In terms of its relationship to normative ethical theory, the

[18] Given the availability of other pragmatist accounts which take a naturalistic approach to morality, one might wonder why I am turning to James for a state-of-nature model of moral status. For example, in *The Ethical Project*, Phillip Kitcher advances a naturalistic story about the emergence of morality, which appeals to the evolutionary advantages that it would have conferred. In particular, Kitcher views basic moral dispositions and the capacity for normative guidance as remedying what he calls "altruism failures"—self-interested tendencies that stand in the way of social cooperation (Kitcher 2011, 103). Another recent ethical framework that is self-consciously pragmatist and naturalistic can be found in Mark Johnson's *Morality for Humans*. Like Kitcher, Johnson views his project as in step with evolutionary psychology, however, his primary aim is to develop an account of moral deliberation that is compatible with and informed by recent work in cognitive psychology and neuroscience (Johnson 2014). Although the Jamesian pragmatic genealogy offered here is a different kind of naturalistic project, it is ultimately complimentary to these other pragmatist attempts to explain morality in evolutionary terms. In particular, pragmatic genealogies can supplement evolutionary accounts by helping them avoid what Nicholas Smyth has called "continuity failure" (Smyth 2017 ,1137). Briefly—and this is especially true of Kitcher's framework—evolutionary accounts of morality tend to assume (often erroneously) that the historically distant conditions under which morality is thought to have served some social function still obtain in the present. While one may be able to legitimately explain the emergence of morality by appealing to the advantages it confers (e.g., promoting social cohesion, resolving disputes), such an explanation is limited to a very specific set of social conditions which, at least arguably, no longer exist. There are two reasons why a pragmatic genealogy can help remedy these continuity failures. First, because they begin with a highly abstracted, generic model of human purposes and needs (instead of, say, those conditions which existed during the Pleistocene), the pragmatic genealogist's state-of-nature model makes a claim about functionality that can be deployed across a much broader range of contexts. Second, by making room for historical de-idealization, pragmatic genealogies are more sensitive to the kinds of changes in social context that may generate continuity failure, and are thereby better positioned to make theoretical changes to the model that accommodate them.

paper has been read as an endorsement of utilitarianism, deontology, or some combination of the two.[19] Others have taken its central focus to be metaethical, stopping short of advocating for any substantive normative claims.[20] Recently, some commentators have attributed to James an anti-theoretical stance, whereby ethics is understood as therapeutic, hortatory, and ultimately concerned with self-transformation rather than normative foundation-building or prescription.[21] In proposing to read "The Moral Philosopher and The Moral Life" as an exercise in pragmatic genealogy, I am less interested in taking a firm position on this debate than in extracting from James's writings a kind of explanatory framework. Nonetheless, I shall try to describe how the reading advanced here leaves room for agnosticism about the question of James's own theoretical ambitions. Thus, I can claim that my view is compatible with either the claim that James does end up advancing substantive normative theoretical claims or with the more therapeutic reading.[22] At the same time, I identify a metaethical thrust in the essay.

[19] For a brief overview of consequentialist versus deontological readings of James' moral thought, see Marchetti (2015, 52).

[20] See, for example Aikin and Hodges (2018). Michael Cantrell has argued that James endorses a kind of metaethical divine command theory (Cantrell 2013, 2). Henry Jackman sees James as developing a kind of "semantic fallibilism" about ethics, which raises the possibility that our moral claims may lack objective truth conditions (Jackman 2019, 2). In their paper, "Three Challenges to Jamesian Ethics", Scott Aikin and Robert Talisse interpret James as offering a bipartite ethical framework, on which a pluralistic theory of value is held to entail *meliorism*—a normative thesis which requires that as many demands be satisfied as possible (Aikin and Talisse 2011). In response to these authors' critique of this framework, Todd Lekan has suggested that James' metaethical defense of James' "inclusivity principle" (which demands that we "create a world that allows for the greatest diversity of ideals and demand satisfaction" [Lekan 2018, 81]) needs to be seen in light of James' relational view of the self and his "experiential approach" (Lekan 2018).

[21] See especially Marchetti (2015).

[22] Marchetti offers some helpful interpretive labels here. On the one hand, he distinguishes the "substantive approach" to James' approach to moral philosophy from the "methodological approach" (Marchetti 2015, 14). Whereas the former regards James' work as addressing specific theoretical problems that still confront us today, the latter picks up on its "form" or "purpose", appreciating his "characteristic way of elucidating certain discourses and problematizing determinate philosophical assumptions" (14). On the other hand, Marchetti distinguishes between the "systematizers" and the "inconclusivists" (15-16): whereas the former "have fervidly refuted the idea of an articulated defense of a moral position detectable in his writings, while… [the latter] – the "systematizers" – have attempted its reconstruction. While the inconclusivists tried to show, in various ways and with different aims, either

Ostensibly, James' aim in the essay is "to show that there is no such thing possible as an ethical philosophy dogmatically made up in advance" (James 1956/1891, 184). He arrives at this conclusion by considering what he takes to be the three fundamental questions of moral philosophy. Before discussing in detail his treatment of these questions, allow me to say how I think they are related.

James' discussion of the psychological question, which "asks after the historical *origin* of our moral ideas and judgments" (James 1956/1891, 185), is mostly negative, or critical.[23] It undercuts what might seem like the most obvious way of providing a naturalistic account of our moral judgments, namely, one which reduces them to associations with experiences of pleasure and pain. James' treatment of the metaphysical question—asking "what the very *meaning* of the words 'good,' 'ill,' and 'obligation' are" (185)—can be read as a constructive attempt to offer a naturalistic, but non-reductionist account of our moral vocabulary. It is here that I think we can read James as advancing a state-of-nature genealogy. Finally, his discussion of the casuistic question, which "asks what is the *measure* of the various goods and ills which men recognize, so that the philosopher may settle the true order of human obligations" (185), draws a set of conclusions from this model. Depending on how one reads him, this is where he argues that moral philosophy—conceived of as a theoretically ambitious enterprise—ought to be either reconstructed or abandoned. For my purposes, it is James' achievement in model-building *via* the metaphysical question that is most significant.

the theoretical weakness or the sporadicity and inconsistency of James's philosophical reflection on ethics, the systematizers have argued in favor of its theoretical solidity and organic articulation" (16).

[23] On this point, I agree with Henry Jackman (Jackman 2019, 7).

Four years before James delivered his address to the Yale Philosophical Club, Friedrich Nietzsche prefaced one of his best-known works by bemoaning the fact that—barring the writings of a handful of benighted Englishmen—the question of the history of our moral values had been largely ignored.[24] Had Nietzsche been in New Haven, he would have likely found in James' discussion of the psychological question more than a faint echo of his own critique of the "English psychologists". Despite praising "the Benthams, and the Mills, and the Bains" for "taking so many of our human ideals and showing how they must have arisen from the association with acts of simple bodily pleasures and reliefs from pain" (James 1956/1891, 186), James is quick to deny their reductive conclusions. Our values, normative reasons, and ideals are just too varied, complex, and contingent to be fully captured or explained "in this simple way" (186). As Henry Jackman puts it, James argues that "[m]oral sentiments are often 'brain-born' preferences coming through the 'back door' of accidental variations, rather than conclusions coming through the front door" of personal, or even species-level, experience" (Jackman 2019, 7).[25] The contents of our evaluative judgments are, therefore, shaped by a wide range of forces which render them irreducible to a small set of psychological relations. This point is essentially a negative "anti-psychologistic" result, which allows James to block a tempting model of "dogmatic philosophy made up in advance", namely, that all of morality (i.e., ideals, judgments, etc.) *just is* simply reducible to associations of pleasure and pain. Given James' value pluralism, this anti-reductionism should not come as a surprise. However, the downside is that undercutting

[24] As Nietzsche puts it: "we need a critique of moral values, the value of these values should itself, for once, be examined – and so we need to know about the conditions and circumstances under which the values grew up, developed and changed (morality as result, as symptom, as mask, as tartuffery, as sickness, as misunderstanding; but also morality as cause, remedy, stimulant, inhibition, poison), since we have neither had this knowledge up till now nor even desired it. People have taken the value of these 'values' as given, as factual, as beyond all questioning" (Nietzsche 2006/1887, 7-8).

[25] Jackman's language here is a reference to the final chapter of James' *Principles* (Jackman 2019, 22 n. 20).

this psychological basis seems to preclude any naturalistic explanation of our moral judgments—indeed, James concludes his discussion by acknowledging that for some, it may signify a victory for the "intuitionist school" (189).[26]

It is in light of this treatment of the psychological question that James' discussion of the metaphysical question should be understood. Given the failure of reductionist naturalism to locate the source of our historically diverse judgments in a single set of psychological relations, James is effectively asking: what can we say about the meaning of our moral concepts? Others have pointed out that an equally felicitous label would be 'the semantic question' (Jackman 2019, 3; Slater 2007, 11). I agree with these assessments, but would insist that James' aim is not exclusively semantic, but also genealogical.[27] That is, he is not simply asking what terms such as 'good,' 'ill,' 'obligation,' *mean*, he wants to know how it is that we came to employ them in the first place (how they came to have their *status* as moral concepts).[28] Like the pragmatic genealogists discussed in the previous section, James' method involves abstraction. He begins by isolating generic features of human life in order to shed light on how concepts such as "obligation," or "claim," would take on their distinctive character. Unlike Craig or Williams, however, James' state-of-nature genealogy is additive, gradually introducing elements until he

[26] While G. E. Moore was not who James had in mind, *Principia Ethica* draws precisely this kind of conclusion (i.e., non-naturalistic intuitionism), from the perceived failure of a naturalistic analysis of moral terms—especially as such an analysis was supposed to involve psychological associations with pleasure and pain.

[27] More specifically, as I shall explain, it bears significant structural similarities to the models one finds in the work of Craig, Williams, and other pragmatic genealogists. It would, therefore, be fair to characterize James' project as a genealogy of semantics. In several places, Marchetti refers to James' project as "genealogical" (Marchetti 2015, 5; 177), but he does not offer a sustained discussion of what this entails.

[28] At several points in the essay, James refers to the *status* of moral terms and relations. For example, he writes that "Surely there is no *status* for good and evil to exist in, in a purely insentient world" (James 1956/1891, 190). And that, once a sentient being is introduced into the universe "[m]oral relations now have their *status*, in that being's consciousness" (190).

derives a satisfying picture in which the target phenomena (i.e., elements of a moral vocabulary) are rendered functional. Indeed, the order in which these elements are introduced is important.

The first stage of his state-of-nature model envisions "an absolutely material world, containing only physical and chemical facts, and existing from eternity without a God, without even an interested spectator" (James 1956/1891, 189). In what can be regarded as a key anticipation of twentieth century constructivism or expressivism, James insists that there would be no values (hence, there could be no moral vocabulary) in a world devoid of *valuers*.[29] As he puts it, "neither moral relations nor the moral law can swing *in vacuo*. Their only habitat can be in a mind which feels them" (190). As a second step, James adds to this model a single sentient creature, contending that only with this addition would value be introduced into the universe. However, although James allows that such a creature could make judgments about what is 'good' or 'bad,' there is an important sense in which he denies that an inhabitant of such a "*moral solitude"* could possess *our* concepts of *good* or *bad*. In part, this is because judgments involving these terms would not be truth-apt. "In such a universe", James explains, "it would of course be absurd to raise the question of whether the solitary thinker's judgments of good and ill are true or not. Truth supposes a standard outside of the thinker to which he must conform; but here the thinker is a sort of divinity, subject to no higher judge" (191).[30]

James' third step is to "introduce a second thinker with his likes and dislikes into the universe" (191). Though necessary, James is unwilling to grant that this addition alone is a sufficient condition for the development of a moral vocabulary. This is, to put it bluntly, because

[29] See Aikin and Hodges (2018) for a compelling case that James can be read as a metaethical expressivist.

[30] As Scott Aikin and Michael Hodges point out, this is why James should not be read as endorsing subjectivism (Aikin and Hodges 2018, 642).

the inhabitants could simply ignore each other. A "moral dualism" or any moral pluriverse, might be replete with values, but it would not yet be a place with *moral* concepts (192). As James explains, "Not only is there no single point of view within it from which the values of things can be unequivocally judged, but there is not even a demand for such a point of view, *since the two thinkers are supposed to be indifferent to each other's thoughts and acts*" (192, italics added). This point is a subtle one, but the idea is that the need for moral concepts emerges only if a fourth addition to the model is made: *psychological capacities which both enable and direct these creatures to recognize and respond to each other's claims*. James does not spend time unpacking this point, but it is clearly intimated in the language that he uses to describe the interplay of claim and obligation. A constitutive feature of *claims*, is that they are "livingly acknowledged" by another consciousness (James 1956/1891, 196). Again, James does not discuss these psychological features in any detail, but presumably, there are a number of general attitudes or capacities which could do this work (e.g., altruism, the desire for approbation, even the self-interested recognition of the benefits of mutually advantageous cooperation).

It is only when these disparate features (i.e., a multitude of sentient beings who possess the capacity to recognize and respond to each other's interests) come together that we get what James calls *a moral life*. As he puts it,

> Wherever such minds exist, with judgments of good and ill, and demands upon one another, there is an ethical world in its essential features. Were all other things, gods and men and starry heavens, blotted out from this universe, and were there left but one rock with two loving souls upon it, that rock would have as thoroughly moral a constitution as any possible world which the eternities and immensities could harbor. It would be a tragic constitution, because the rock's inhabitants would die. But while they lived, there would be real good things and real bad things in the universe; there would be obligations, claims, and expectations; obediences, refusals, and disappointments; compunctions and longings for harmony to come again, and inward peace of conscience when it was restored; there would, in short, be a moral life, whose active energy would have no limit

> but the intensity of interest in each other with which the hero and heroine might be endowed (197).

It is against this background, for James, that our moral concepts begin to make sense. There would be no *use* for the notions such as 'obligation' given the conditions of moral solitude; nor would there be in a world in which sentient beings lacked the psychological compulsions to take notice of each other. Moral differences, in such cases, would not be ones that made a difference.

I propose that we can think of James' notion of the *moral life* as a moral state-of-nature. It serves as a model incorporating basic (and highly generic) features of human life which enable certain features of our conceptual practices to be understood as *practical eventualities.* I shall return to this model to motivate the claim that *moral status* can also be understood as a practical eventuality, but first I should digress briefly to touch on James' third question, as this will round out my interpretation of his view.

The move from the metaphysical to the casuistic question involves a transition from what I am calling James' state-of-nature model to what can be reckoned as the de-idealized present. The former discussion attempts to explain on the basis of generic factors the emergence of a moral vocabulary, whereas the latter introduces features endemic to our local moral practices: a world replete with competing values, in which we cannot agree about God's existence, in which we encounter and need to take into consideration past attempts at moral system-building, and where new proposals for evaluative unification are always being thought up—often from unexpected sources.[31] The casuistic question considers the possibility of imposing an ordering on

[31] James writes, "The last fundamental question in Ethics was, it will be remembered, the *casuistic* question. Here we are, in a world where the existence of a divine thinker has been and perhaps always will be doubted by some of the lookers-on, and where, in spite of the presence of a large number of ideals in which human beings agree, there are a mass of others about which no general consensus obtains" (198). In addition, James recognizes that our present situation is replete with different theories of the good (200)—none of which can plausibly be said to command universal acceptance.

these competing values—especially as the task is undertaken by the figure of the moral philosopher.[32] If the metaphysical question can be read as addressing the etiology of our moral vocabulary, the casuistic question can be read as an etiology of something like *moral philosophy* conceived as a theoretical enterprise aimed at discovering "the *measure* of the various goods and ills which men recognize, so that the philosopher may settle the true order of human obligation"(185). Another way of putting this is to say that James is here depicting moral theory as a practical eventuality.[33] One conclusion which I think is safe to draw from this discussion is that James takes (i) the sketch of moral life gleaned from the metaphysical question, along with (ii) his claim that our current world is "tragically practical" (202), to undermine a traditional sort of foundationalism in moral philosophy (James 1956/1891, 209). For example, he casts serious doubt on the project—common to so many figures—of theorizing an abstract moral order which stands behind our practices, which could be used to evaluate them. My own view is that beyond this rejection of foundationalism, what James' own position is remains a matter that is subject to

[32] James attributes to this figure a number of commitments: (i) an unwavering rejection of moral skepticism (184); (ii) a commitment to developing a unified and stable moral system (185); (iii) a commitment to objectivity, which implies refraining from partisanship with respect to any given ideal (203, 204). The extent to which James "identifies" with this figure is, I think, *the* central interpretive question. One feature of "The Moral Philosopher and the Moral Life" that lends support to my "genealogical" reading is that the figure of the moral philosopher does not enter into James' discussion of the psychological or the metaphysical questions. If I am right, this is because systematic moral philosophy depends on a set of historical conditions which are only introduced in James' discussion of the casuistic question.

[33] This is not to say that James is necessarily *vindicating* moral philosophy's methods and ambitions. Rather, it is an attempt to explain why a certain set of problems would have taken the shape that they do. Somebody—an extra-terrestrial "anthropologist", perhaps—might wonder what it is that drives human beings to engage in the peculiar enterprise known as moral philosophy. James' answer to this query, as I understand it, combines the genetic story offered in response to the metaphysical question with the thought that (as the result of historical chance) the modern world is overflowing with diverse values, competing projects, and incommensurable worldviews. These conditions give rise to certain kinds of practical and theoretical questions, which James' moral philosopher is committed to answering in a particular way. Whether one ought to interpret James as having regarded such problems as legitimate and worthy of continued engagement turns, therefore, on the extent to which one is prepared to regard him as identifying with the moral philosopher.

reasonable disagreement. But since the remainder of my discussion does not hang on this fascinating interpretive question, I shall pass it over.

I have been urging that James's "moral life" describes a state-of-nature model which explains how our moral vocabulary came about. In light of the fact that we are the kinds of creatures capable of taking up evaluative attitudes, who inhabit a world with other valuers, who communicate, and who are at least minimally disposed to care about the claims of others (whether this be explained in terms of a desire for harmony or some other psychological mechanisms), it makes sense that we have come to talk of *values, goods, obligations,* and other moral notions. Moreover, given that we find ourselves in the "tragically practical" moral present, it makes sense that we are the kinds of creatures who would engage in certain kinds of inquiry about value—namely the "casuistic" attempt at imposing order on our values. In this respect, James's model is able to "retrodict" our propensity to engage in certain kinds of philosophical disagreements.

More importantly, however, I believe that James's model also sheds light on how the concept of moral status could have arisen. One can anticipate circumstances in which participants in the moral life would need some way of identifying appropriate targets of moral concern and of generalizing about their obligations. These circumstances, which can be understood as distinctive kinds of interpersonal disagreements, serve as practical pressures for a proto-concept of moral status to emerge. In other words, James' account can explain not only the emergence of our moral vocabulary, but it also explains why we would come to engage in disagreements about *who* or *what* counts as an appropriate locus of moral consideration. Allow me to unpack this in more detail.

6.2.3 Moral Status as Practical Eventuality: Moral Interpolation, Deontic Generalization, and Appropriate Loci of Moral Concern

To substantiate the claim that a proto-concept of moral status would be a practical eventuality within a moral state-of-nature, I must identify a set of generic needs to which it can be seen as a reasonable response.[34] My contention is that these needs can best be understood as stemming from a phenomenon that I shall call *moral interpolation*. Put very roughly, moral interpolation denotes a process through which moral practices are extended or transformed in light of novel circumstances. It can be considered a form of problem-solving requiring the ability to determine morally appropriate responses to new situations, especially in the absence of explicit rules or instruction. My central claim is that *as forms of moral interpolation develop in complexity and sophistication, there arises a distinctive set of needs which the proto-concept of moral status would satisfy.* In particular, I shall argue that in its generic, rudimentary form endemic to the moral state-of-nature, moral interpolation generates quasi-inevitable forms of interpersonal disagreement concerning the scope of an individual or community's moral considerations. Although I believe that the target conceptual practices (i.e., involving the proto-concept of moral status) could be expected to help *resolve* these disagreements, my claim is more that the proto-concept would enable them to be *expressed* as disagreements. Moral interpolation requires (i) that agents are capable of identifying *possible loci* of moral consideration; and, (ii) that they are able to make generalizations about their obligations. It is by allowing agents to satisfy these practical requirements that the proto-concept of moral status makes its primary

[34] These needs should be analogous to the need to gather and pool information to which both Williams and Price appeal in their accounts, or the need to flag good informants to which Craig appeals.

functional contributions to the sort of generic moral practices that the state-of-nature model describes.

Moral interpolation involves a practical ability to determine morally appropriate forms of response without explicit instruction. As I hope will become clear, this term is meant to capture a broad range of phenomena of varying sophistication and complexity. For example, while interpolation can involve conscious deliberation, it is often done tacitly and pre-reflectively. In so far as it involves (conscious) deliberation, moral interpolation can be undertaken individually or collectively. It may involve cognitive, affective, and conative aspects, as well as the imagination. Although I shall argue that moral interpolation is an inevitable feature of moral practice, it can come to be expressed in different ways. That is, although I am concerned here with indicating the place of moral interpolation within an abstracted model, one would need to look to history to understand how it comes to be elaborated. I shall return to this point in Section 6.3 when I discuss how deliberating about practical identities and articulating conceptions of moral progress constitute sophisticated forms of moral interpolation that have histories.

Even the inhabitants of James' sparsely populated rock would require a set of practical capacities enabling them to extend their practices into the future. The moral life requires its participants to be able to reason about hypothetical cases, to integrate new candidates for moral concern and their claims into the picture, and to be capable of projecting their obligations onto future cases in light of ambiguity or unforeseen circumstances. Moral interpolation denotes this array of often tacit practical abilities.

The simplest cases of moral interpolation can be regarded as a byproduct of the holistic, inferential nature of obligations (or, more generally, any kind of normative engagement). Obligations do not wear their applications on their sleeve but require a range of capacities including, but not limited to, sensitivity to context, an often tacit understanding of their implications, as well as an awareness of fairly complex counterfactual considerations. In order to grasp, for example, that one has an obligation to keep one's agreement to weed the communal garden, one needs to have a sense of the sorts of considerations which might nullify the agreement (the presence of poisonous snakes, for example), that obligation extends into the future (however indeterminately), that if one were unsure as to whether *this* plant is a weed, one should ask before one pulls, and so on. In this respect, carrying out one's obligations is analogous to applying a concept. Just as properly deploying the concept *red* requires a certain level of mastery of a range of material inferences, so too does acting to fulfill an obligation require a considerable amount of background knowledge and practical competence. If, in Robert Brandom's terms, "*discursive* activity…is inseparable from the inferential activity of giving and asking for *reasons*" (Brandom 2009, 8), then I am suggesting that something like "deontic activity" is inseparable from inferential activity (broadly construed) as well. What makes grasping and fulfilling obligations a matter of interpolation, is that doing so involves an extension of aspects of a moral practices—determining appropriate responses, forming and adjusting attitudes, adopting patterns of concern in novel situations—without reliance on explicit rules. There is, I am claiming, a kind of practical inertia operating here. An inertia that is constitutive of any conceptual practice and which demands certain kinds of competencies or practical abilities of their participants. This is important because, it allows me to claim that interpolation—at least in this very basic sense—is a generic feature of human practices, and thus, is a candidate for belonging to the moral state-of-nature.

Of course, agents need not consider any of this explicitly. Moral interpolation can form part of the background norms and dispositions that are constitutive of a practice, and is often a matter of implicit training or habituation. That being said, it *can* be (and often is) subject to disagreement and criticism. For my purposes, it is crucial to note that interpolating, in these simple cases I have been discussing, will often presuppose some understanding of the kind of thing to which the moral concern is directed. For example, part of what is required for me to properly grasp my obligation to tend your plants while you are away, is an understanding of the kinds of conditions that plants require to survive. In cases of direct obligations, moreover, the requirement is that one grasps that the "target" of such an obligation is the kind of thing that requires certain kinds of treatment and consideration. But even more fundamentally, there is a sense in which grasping an obligation demands a prior determination that the target is an appropriate locus of (an indefinite range of) practical and moral concern in the first place. That is to say, at a very basic level, the kind of moral interpolation constitutive of grasping obligations rests on a practical ability to distinguish between appropriate and inappropriate targets of moral and practical concern. If this is right, it suggests that, even in the simplest cases which I have been discussing, there is space for a proto-concept of moral status. Though it may not be made explicit, moral interpolation presupposes an ability to identify appropriate loci of moral concern.

But there are more sophisticated forms of interpolation which require not just an implicit practical ability, but an explicit vocabulary with which to talk about (i.e., disagree about, deliberate about) what counts as an appropriate target of moral concern. Consider, for example, cases of moral interpolation involving group obligations. Jones is no longer concerned simply with his obligations to Smith, but to his obligations towards members of a social group—that is, he comes to consider claims made on behalf of a "we". This seemingly inevitable, though arguably more complex form of moral interpolation requires that agents are able to identify appropriate candidates for group membership. This ability may turn out to be a trivial one, in that the scope of the “we” is simply obvious to everyone, and hence, does not require further reflection (be it on Jones’ or anyone else’s part). However, matters may be indeterminate, in which case there may turn out to be disagreement—perhaps requiring collective deliberation—about how to determine the “we’s” extension. In either case (as in the simpler form of interpolation in response to holistic considerations) an implicit practical ability is being presupposed: the capacity to pick out appropriate targets of practical concern. But when a certain level of indeterminacy and disagreement is allowed, it is plausible that there would emerge a need for a vocabulary enabling speakers to identify potential loci for moral engagement. Moreover, this need would become more pressing as disagreements become more extreme, for example, when questions are raised about potential claimants who fall outside the immediate community—such as strangers, non-human animals, or even sacred objects. Thus, in light of these increasingly sophisticated forms of moral interpolation, a proto-concept of moral status becomes a practical eventuality in so far as it enables speakers to identify and deliberate about what counts as an appropriate loci of moral concern. That is, although basic forms of moral interpolation require a practical ability to identify appropriate loci; as these practices gain in complexity there are pressures on the community to develop a vocabulary that can render these practical abilities explicit, subject them to various forms of assessment.

In addition to this target-identifying function, another reason that the proto-concept of moral status could be considered a practical eventuality has to do with the need for generalization about moral considerations. It follows from the holistic, inferential features that obligations, that agents may be driven to reflect upon, or in some sense refer to their obligations *en masse*—to be able to make *deontic generalizations,* as I shall call them. The proto-concept of moral status can be thought of as meeting this need.

On the one hand, speakers must be able to express that different kinds of moral considerations correspond to different kinds of things. For example, that certain kinds of moral responses may be appropriate for adults, but not for small children (or sacred objects, but not unhallowed ones). Drawing these distinctions, in turn, requires being able to make interpolative generalizations or the capacity to "range over" the sorts of practical considerations appropriate to different cases. Again, although in many instances this may be a tacit, practical ability that does not find expression in a language; in other cases—especially when disagreement arises—there would be practical pressure for an explicit vocabulary. On the other hand, the ability to make these kinds of generalization is also a prerequisite for determining comparative obligations, or in weighting them. Indeed, James himself seems to allow for these cases by acknowledging the existence of "insignificant persons" (195). An insignificant person is someone for whom obligations *in general* are afforded less weight. The point is that the very idea of recognizing comparative significance presupposes the ability to make generalizations about one's obligations.[35]

I can now bring together the various threads of my discussion of moral interpolation by returning to the question of how James' moral state-of-nature allows for a functional explanation of a proto-concept of moral status. In Section 6.2.1, I claimed that state-of-nature models make use of a kind of functional explanation taking the following form:

Target phenomenon P performs function F given circumstances C.

[35] There is an analogy here with disquotational accounts of truth. As I mentioned in Chapter Two, instead of proposing a substantive theory about the nature of truth, semantic deflationists often point to the fact that the term serves as a helpful grammatical device which allows speakers to make generalizations about a range of propositions to which they want to commit themselves, but where doing so would be impossible or redundant. My claim is that a proto-concept of moral status would perform a similar grammatical role enabling agents to specify a range of commitments.

In this case, the target phenomenon, *proto moral status*, performs the function of enabling agents to engage in moral interpolation. It does so, primarily, by allowing speakers to identify appropriate loci of moral concern and to make generalizations about their moral obligations. It can be understood as discharging this function in light of a set of circumstances which require agents to modify or extend their practices, especially when internal disagreement is to be expected. For example, when questions arise about the extension of group membership or when certain kinds of practical concern are reserved for a subset of some group.

So far, these initial hypotheses are simply meant to provide the basis for a functional and etiological account of our conceptual practices involving moral status. They establish a set of generic needs to which those practices could be expected to answer. Before moving on to show how a de-idealized state-of-nature model can yield functional diversity, allow me to clarify a few details of this proposal and to ward off some potential objections.

6.2.4 Moral Interpolation: Further Thoughts

First, in the previous chapter I mentioned that both David DeGrazia and Christine Korsgaard have suggested that the concept of moral status (i) operates as a variable or placeholder; and, (ii) allows agents to generalize about their obligations (DeGrazia 2008, 184; Korsgaard, 2018, 96). Given that I claimed that these proposals required further development, I ought to say how the present account differs from them. So far, the state-of-nature genealogy's main innovation lies in the fact that it shows how *target identification* and *generalization* support moral interpolation. Instead of simply asserting that the concept of moral status operates as a variable or allows for deontic generalizations, I have tried to depict these functions as responses to generic needs arising out of a central feature of moral practice—namely, moral interpolation. Not only does this begin to bridge the explanatory gap left open by DeGrazia and Korsgaard's accounts (i.e., it explains why, rather than assumes that moral agents would need a concept to generalize about their obligations or serve as a placeholder), in doing so it can also avoid begging the question against the eliminativist. More importantly, in connecting these basic functions to the phenomenon of moral interpolation, my account opens up space for further inquiry. Indeed, in the next section, I will investigate how these rudimentary functions (i.e., target identification and generalization) could be expected to become invested in more complex deliberative practices. In the next chapter, I add even more complexity, of an actual, historical kind. As I shall ultimately argue, it is by looking to these developments that one can gain a better sense of how the concept of moral status would acquire functional diversity, thereby allowing for a response to the problem of intractability.

This potential for historical elaboration raises a second set of possible questions regarding my claim that the proto-concept's eventuality involves the making explicit of some implicit abilities. It might be asked: Is not the emergence of the concept of moral status itself an historical question and not something that can be settled by a priori reflection? And is it not possible that in a state-of-nature, the need for an explicit concept would never arise?

I ultimately agree that the historical question is important. To understand fully the role that moral status plays in our conceptual practices one needs to look to how those practices have developed. However, the aim of the state-of-nature model (at least up to this point) is not historical understanding but, rather, something like conceptual understanding. In this respect, my explanation bears a family resemblance to Robert Brandom's expressivism about logical terms. For Brandom, a logical vocabulary is neither inscribed in our souls by God or Nature, nor written into the fabric of Reality. Rather, a logical vocabulary allows humans to express something that they initially knew how to do implicitly. In particular, it allows language-users to make "explicit the inferences that are implicit in the conceptual contents of nonlogical concepts" (Brandom 2000, 61). Consider the purpose of the conditional. Brandom writes:

> Prior to the introduction of such a conditional locution, one could *do* something, one could treat a judgment as having a certain content (implicitly attribute that content to it) by endorsing various inferences involving it and rejecting others. After conditional locutions have been introduced, one can *say*, as part of the content of a claim…*that* a certain inference is acceptable. One is able to make explicit material inferential relations between an antecedent or premise and a consequent or conclusion (Brandom 2000, 60).

In making this point, Brandom's aim is, I take it, to shed light on our conceptual practices (in this case, logical ones) by appealing to their expressive functions. Doing so, however, is perfectly compatible with an historical investigation. In a similar vein, just as Brandom's expressivist insight could be paired with an historical investigation into logical vocabularies, so too, a state-of-nature model explaining how creatures like us began to deploy the vocabulary of moral status is entirely compatible with an historical look at different conceptions of moral status.

Still, one might press the point further: couldn't there be a world in which we *never* acquired the explicit vocabulary with which to attribute moral status? Even if one concedes the need for the implicit ability, a critic might ask: do we really need the vocabulary to express it explicitly? The eliminativist, of course, will be especially eager to raise this point as it would suggest that the concept of moral status is one with which we could afford to dispose.

My hypothesis is—as I explained in the previous section—that the need for an explicit moral status concept is a function of the degree of disagreement within a linguistic community. The greater the disagreement about the scope of their obligations, the more likely it is that speakers will need the proto-concept. It is, of course, entirely possible that human beings might never have come to develop the kinds of moral practices in which these sorts of practical conflicts mattered, or where they settled such questions by an appeal to some alternative authoritative process rather than through reasoned disagreement. One might imagine something along the lines of Williams' "hyper-traditional society", which is characterized by minimal reflection and disagreement (Williams 1985, 142). Such a society, one might think, would not need an explicit concept of moral status (even if they still would require the practical abilities discussed above). This possibility might be thought to lend some weight to the charge that the concept is less practically necessary than I have made it out to be.

There is both a direct and an indirect line of response to this concern. On the one hand, a direct response might try to show that the very idea of a hypertraditional society is implausible. According to evolutionary psychologists, humans developed a set of on-board psychological mechanisms for drawing in-group/out-group distinctions (Buchannan and Powell 2016). These, one might think, are prone to produce the kinds of disagreements necessary to get my genealogical story off the ground. On the other hand, there may simply be no issue with embracing the possibility: one might accept that the need for an explicit moral status concept is a relatively late historical invention. The eliminativist's original concern was that the concept is redundant *from the standpoint of present practices*. The argument that I am advancing premises the concept's functionality, at least in part, on the quasi-inevitability of certain kinds of disagreement. All that is needed to meet the eliminativist's challenge is to show that disagreement is, at least to some extent, a feature of *our* moral practices which we could not simply wish away. And this I take to be uncontroversial.

A third and final point of clarification concerns the argumentative strategy which I am employing. A central claim of this chapter is that one can derive (from this initial state-of-nature model) a set of more diverse and complex practices in which the concept of moral status has been taken up. In the next section, I shall argue that this functional diversity explains the divergence between individualist approaches to moral status and their humanist critics. One might worry, however, that there is something self-defeating about this strategy, in so far as appealing to the functional diversity of moral status threatens to undermine the concept's unity. If, as I shall ultimately maintain, the concept of moral status has come to be deployed in a variety of different domains and for a plurality of purposes, then why think that we need a single concept to link these diverse uses together? Again, this is a potential concern that an eliminativist is likely to raise.

A similar concern has been voiced with respect to the concept of *personhood*. It is evident that there are many practical contexts in which we are concerned with persons.[36] For example, personhood plays an important role in our evaluations of moral responsibility, prudential rationality, our concern with biological continuity, and our assessments of character. Given the diversity of these practical concerns, some writers have argued that the very attempt to provide a unified conception of *personhood* (e.g., to provide a metaphysical account of personal identity) is misguided (Shoemaker 2007). Analogously, the claim I shall advance in the next section—i.e., that the concept of moral status ultimately comes to play a role in a wide range of deliberative contexts—may provoke the same kind of skepticism.

In her recent book *Staying Alive*, Marya Schechtman provides a response to this kind of worry in relation to the concept of personhood, which, I believe, can be appropriated and repurposed to show that we do need something like a unified concept of moral status which is

[36] Judith Jarvis Thomson has raised a similar criticism of the supposed right to privacy. Her claim is that reflection on what this right is supposed to involve ends up revealing "a cluster of rights—a cluster with disputed boundaries—such that most people think that to violate at least any of the rights in the core of the cluster is to violate the right to privacy" (Thomson 1975, 312-3. But, Thomson thinks, it is far from clear what is supposed to link these various rights together, aside from "their being rights such that to violate them is to violate the right to privacy" (313).

capable of functioning across a range of deliberative practices. A key premise of Schetchman's *person life view* is that we experience persons as unified targets which encompass, but are never ultimately reducible to, a single category of assessment, be it biological, prudential, psychological, or character-based. And although these may be isolated for particular purposes (e.g., my doctor may treat me as a biological entity while running some tests), they all operate together to form part of a unified, person life. In response to the critic who denies that our concept of personhood really does involve this element of unity, Schechtman would have us reflect on our interpersonal relationships. On her view,

> [t]he son I feed and clothe and comfort is the same person I chastise for behaving badly to his sister and the same person to whom I try to teach the value of hard work and explain the benefit of making small sacrifices now for larger benefits later. He is also the same person whose straight As bring me pride and whose disappointments are a cause for my sadness, and the person whose health I am concerned to safeguard. I do not have a moral son and an animal son and a psychological son—I have a single son who has all these aspects and is important to me in all of these ways (Schechtman 2014, 83).

A central function of our concept of "person" is that it allows us to tie together the broad range of practical concerns that we have with individuals, even when those concerns are not co-present. As she puts it, "If we are to acknowledge the fact that different practical relations do not always occur together but also accept that they are not entirely independent of one another we seem to need something like a forensic unit which provides a unified target of our various practical questions and considerations but within which not all of the particular practical relations need apply simultaneously" (63). The concept of *moral status*, I want to suggest, performs an analogous function. To attribute moral status to an entity is to regard it as a unified target of diverse kinds of practical concern. To say that someone has moral status may include, for instance, taking their interests into account when deliberating about what to do, engaging empathetically and imaginatively with their perspective, affording them various kinds of rights and protections, or caring about them in myriad ways. Although these practical concerns differ in important ways, we do tend to think of their targets as being unified in the sense to which Schechtman appeals. The concept of moral status—like the concept of *personhood*—can help make this unity explicit.

So far I have offered an explanation of how our conceptual practices involving moral status could be thought to have arisen. Drawing on James' notion of a "moral life" I constructed a state-of-nature model involving a set of generic human capacities and concerns. I introduced the phenomenon of moral interpolation as a central feature of this model, and argued that it generates a set of needs (i.e., target identification, deontic generalization) to which the concept of moral status can be seen as an apt response. I then fleshed out this proposal by considering some potential objections to it.

In working out a state-of-nature model, I have alluded to some ways in which we begin to see a possible set of responses to the problems of intractability and eliminativism. To employ Kusch and McKenna's helpful terminology, one can already begin to "retrodict" some of the theoretical controversies involving the concept of moral status based on the explanatory model developed so far. For example, I have already suggested how appealing to the basic functions that the proto-concept of moral status performs can provide the basis for a response to the eliminativist's challenge. The claim is not simply that the concept allows speakers to pick out unified targets of practical concern or make generalizations about their obligations, but rather, that these functions answer to a set of practical needs. In the following section I will expand on this response to the eliminativist by introducing a further set of functions that the concept could be expected to play as our practices develop in complexity and sophistication.

6.3 De-Idealization: Moral Status, Practical Identities, and Moral Progress

So far I have laid the groundwork for a state-of-nature genealogy of the concept of moral status, proposing that the concept allows agents to identify loci of moral concern, to make deontic generalizations, and that these functions answer to a set of basic needs and limitations given that we are creatures who engage in moral practices. In this section, I proceed to the second, de-idealizing stage of a pragmatic genealogy. As I mentioned in Chapter Four, de-idealization can occur in one of two ways. To borrow terminology from Matthieu Queloz, a *primary elaboration* aims to show how a proto-practice's "development [is] driven by the practical pressures internal to the model, such as the foreseeable problems which the original solution offered by the proto-practice will bring in its wake" (Queloz 2019, 9). By contrast, a *secondary elaboration* will trace a practice's "development driven by the introduction of increasingly socio-historical local needs into the model and the new problems that come with them" (9). In this section, I shall suggest two ways in which moral interpolation could be expected to have developed, each corresponding to a different sort of elaboration. First, I show how the model could be expected to undergo a primary elaboration as the proto-concept of moral status gets taken up by practices in which agents come to deliberate about or form their practical identities. Second, I offer a secondary elaboration of the model which highlights its connection to recent conceptions of moral progress. Both elaborations support the two theses that I identified at the beginning of this chapter: functional plurality and theoretical correspondence. Together, these claims form the main thrust of my response to the problem of intractability and add further support for a counterargument against the eliminativist.

6.3.1 Moral Status and Constructing Practical Identity

In introducing the notion of moral interpolation, I suggested that the term covers a broad range of phenomena which vary in complexity. It includes, for instance, the kind of tacit, practical knowledge that is presupposed in fulfilling an obligation, as well as the more complex deliberative practices by which moral agents reflect on the meaning and scope of group obligations. Another form of moral interpolation involves the formation and reconstruction of practical identities. Following Christine Korsgaard, a practical identity can be thought of as "a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking" (Korsgaard 1996, 101).[37] I am going to argue (1) that there are different ways of constructing and contesting practical identities. (2) That these differences constitute distinct deliberative contexts within which the concept of moral status gets extended. (3) That one can explain (or at least begin to explain), on the basis of these different deliberative contexts, the theoretical divide between individualists and humanists.

Most of us inhabit multiple practical identities. Examples include one's place within a family, one's national, ethnic, or religious identity, one's profession, memberships to communities, clubs or associations, and pretty much anything that one could regard as "a role with a point" (Korsgaard 2009, 21). These practical self-conceptions "govern our choice of actions, for to value yourself in a certain role or under a certain description is at the same time to find it worthwhile to do certain acts for the sake of certain ends, and impossible, even unthinkable, to do others" (20). It is in this sense that our practical identities can be understood as the source of our reasons, obligations, and integrity.

[37] The notion of a practical identity plays a central role in Christine Korsgaard's constructivist theory of normativity and especially in her more recent account of agency that compliments it (Korsgaard 1996; 2009).

Nearly all of one's practical identities are obviously contingent. Just as nobody chooses their biological parents, nobody decides in advance the cultural, geographical, and historical contexts within which they are socialized. Korsgaard recognizes this but thinks that when we choose to act in accordance with the dictates of a particular practical identity we "make it our own" thus constituting ourselves (Korsgaard 2009, 43). This is because "whenever I act in accordance with these roles and identities, whenever I allow them to govern my will, I endorse them, I embrace them, I affirm once again that I am them" (43).

One upshot that Korsgaard is particularly eager to draw from all of this is that the apparent unavoidability of valuing ourselves under *some* practical identity yields a transcendental argument for the claim that we are committed to valuing humanity as such (Korsgaard 2009, 24).[38] This Kantian conclusion is, of course, contentious—even amongst metaethical constructivists (Street 2012). Fortunately, for my purposes I can remain agnostic about it, and focus instead on Korsgaard's compelling thesis that practical identities serve not only as a source of our reasons but also as the objects of deliberation and critique (Korsgaard 2009, 21).

Practical identities are the sorts of things that admit of interpretation and negotiation. Two aspiring Christians can differ in what they take these identities to require of them (not only that, they can disagree about *to whom their identities obligate them*). Moreover, since any given person is likely to take on multiple identities within their lifetime, conflicts are all but inevitable.

[38] Korsgaard offers the following succinct gloss on this argument: "in valuing ourselves as the bearers of contingent practical identities, knowing, as we do, that these identities are contingent, we are also valuing ourselves as rational beings. For by doing that we are endorsing a reason that arises from our rational nature—namely, our need to have reasons. And as I've just said, to endorse the reasons that arise from a certain practical identity just is to value yourself as the bearer of that form of identity. We owe it to ourselves, to our own humanity, to find some roles that we can fill with integrity and dedication. But in acknowledging that, we commit ourselves to the value of our humanity just as such" (Korsgaard 2009, 24-25).

It is these *deliberative contexts* surrounding our practical identities with which I am concerned. In particular, I want to view them as important cases of moral interpolation, through which our practices get extended, transformed, or modified in ways that do not depend on the application of explicit rules or criteria. Above all, I aim to examine how these different contexts can be understood as calling for different uses of the concept of moral status. Although it is not always at the forefront of her analysis, I take it that Korsgaard would agree that one's attributions of moral status are a function of the practical identities one adopts.[39] Examining the different deliberative strategies by which agents arrive at practical identities, is, therefore, also to examine different deliberative strategies by which agents arrive at moral status attributions. It is by understanding these differences, I shall claim, that a resolution to the problem of intractability begins to emerge.

Consider first what might be called contexts of *identity articulation*. These occur when some existing practical identity is expanded upon or rethought in some fundamental way, or where its implications get worked out through a process of reflection. The example of the two aspiring Christians, trying to determine just what sort of conduct their religious convictions require of them, illustrates what I mean here. As Korsgaard notes,

> there is room for argument about whether a particular way of acting is the best way or the only way to go about being, say, a teacher or a citizen… it is with reference to the role or point of that form of identity that thought and argument about different interpretations of that form of identity can go on. There is room for creativity here, as well as argument: one might find a new way of being a friend (Korsgaard 2009, 21).

It is this room for interpretation and creativity constitutive of identity articulation that makes it a form of moral interpolation. To a certain extent, moreover, practical identities just are continuing

[39] As I mention in Chapter Five, Korsgaard's clearest statements of her views on moral status (she used the term "moral standing") occur in her recent book, *Fellow Creatures*.

projects of collective identity articulations. For my purposes, it is crucial to recognize that identity articulation can be a means of arriving at moral status attribution. At least in some cases, part of what is involved in, say, figuring out what it means to be a Christian, is to arrive at a determination about to *whom* or to *what* it is appropriate (or perhaps, required) for you to display certain forms of moral consideration. Some forms of identity articulation will involve more determinate claims about the scope of those obligations that the identity entails. Others may not make such sharp determinations, leaving it unclear as to the scope of moral or practical concern that they demand. The point, however, is that identity articulation can be, and often is, a central and salient deliberative context within which commitments about moral status take shape.

A second deliberative context occurs as the result of a conflict between distinct practical identities. *Identity conflict* comes in both an external and an internal form. Taken externally, identity conflicts are just special forms of inter-personal disagreements, as exemplified in, say, the conflict between the pro-life and pro-choice activists. But often enough, identity conflict is internal, the result of our various self-conceptions pulling in different directions. Indeed, in "Justice as a Larger Loyalty" Richard Rorty goes as far as to claim that internal identity conflicts just are the source of moral dilemmas (Rorty 2007, 45).[40] It is important to note—indeed part of what I shall argue hangs on this—that identity conflicts will appear the most pronounced when there is little or no overlap between them, or between the kinds of actions or commitments that

[40] There is an interesting (and arguably, longstanding) question about the extent to which these kinds of conflicts ought to be resolved. There is a robust tradition in political philosophy—going back to Book IV of Plato's *Republic* which sees a kind of psychological unification or "harmony of the soul" as a necessary component of human flourishing, and even of effective political participation. This view is shared by Aristotle, Rousseau, and Kant. Rorty's insistence on the separability of projects of private self-perfection from public projects of social concern puts him in opposition to this trend (Rorty 1989). In doing so, I take him to be in the company of something like a "counter-tradition" that goes back to at least Machiavelli, and which includes Hobbes, Hume, and Nietzsche. Korsgaard places her own project squarely within this unificationist tradition, whereas Rorty arguably leaves room for a kind of disunity, at least when it comes to the separation of our private projects of self-creation from our public responsibilities to others.

they entail. The most irresolvable-seeming moral disputes are those which result from a clash between mutually exclusive practical identities where there is no further shared identity to which disputants can readily appeal.[41] However understood, for my purposes, the important point is that these forms of disagreement also create sites for disagreement about moral status.

A third kind of case involves *identity creation*. To a certain extent, the creation of a new practical identity is bound up with the emergence of new social roles and the institutions of which they are a part. Identity creation is probably best described as a social endeavor, rather than as something that individuals undertake on their own—if only because any practical identity will require some degree of social recognition. If the reasons and obligations derived from a private identity were, in principle, unintelligible or incommunicable to others, this would belie their normative status as obligations.[42] Moreover, new identities probably do not come into being fully-articulated, but will require a fair amount of elaboration and negotiation. Just like the previous two deliberative contexts, identity creation can shape profoundly the manner in which agents come to attribute moral status.

At the risk of oversimplify things, I want to distinguish between two processes of identity creation which have important implications for thinking about moral status. On the one hand, the formation of a practical identity might begin with a specification or "core description" of a role,

[41] Kantian constructivists like Korsgaard will, of course, insist that there always is a shared practical identity (without which disagreement would be, at least in a constitutive sense, impossible) namely, the shared practical identity of creatures endowed with reflective consciousness. Humean constructivists, such as Street, do not find this argument persuasive. They argue that it presupposes a theoretical standpoint wherein an agent is required to shed her normative commitments and, subsequently, to ask whether she has a reason to endorse normative reasons in general. The problem, according to Street, is that this kind of question cannot coherently be raised by a constructivist, for whom there are no values or normative reasons which float free of some evaluative stance (Street 2012).

[42] Korsgaard (1996) argues, *via* Wittgenstein's considerations against the privacy of meaning, that reasons must be public (at least in principle). Given that she views our reasons as stemming from our practical identities, it is tempting to conclude that Korsgaard would find the idea of a "logically private identity" (in the sense of an identity which could not be communicated) just as unintelligible as an incommunicable, private sensation.

and then proceed to work out the kinds of obligations that go along with this role. Call this first possibility the *thick-to-thin model* of identity creation. On the other hand, some practical identities appear to take shape in the opposite way: beginning with a specification of some core obligations, a task which is then followed by an elaboration or a description of some shared social role. Here, the process is *thin-to-thick* in the sense that obligations precede a broader social meaning attached to the identity.

As an example of the first kind of identity creation, consider the emergence of new social positions such as "the data scientist." Initially, one might think, this new practical identity was formed in response to novel scientific and technological problems. Conferences were held, journals formed, and certain institutional roles were created, and thus there emerged a new kind of "role with a point." Initially, the practical identity of the data scientist may not obviously have been connected to a distinctive set of ethical concerns or obligations. But as the broader social implications of data science became clear to a greater number of people, the practical identity underwent a kind of articulation, through which a set of distinctively ethical obligations and responsibilities were attached to the role. By contrast, an example of the second—"thin-to-thick"—kind of identity creation might be the emergence of ethical veganism as a moral identity distinct from dietary (or even ethical) vegetarianism. In 1944, The Vegan Society was formed after a protracted schism within the British Vegetarian Society (Wrenn 2019, 190). The former association—which popularized the term 'vegan' and has influenced significantly veganism as a social movement—was, at least in its inception, in a position in which it needed to demarcate itself from vegetarianism. Its initial mission statement reflects primarily a set of practical concerns, such as dietary rules, promoting the rights of non-human animals (as well as vegans). In this respect, veganism as a practical identity was initially centered around a set of distinctively moral obligations. As Corey Lee Wrenn has suggested, veganism as a social movement has grown and transformed in a multitude of ways since the 1940s (Wrenn 2019). As it has done so, a great deal of reflection about the broader significance of the veganism as a practical identity has taken place. Though initially formed—at least primarily—around a set of obligations, it's social meaning has been and continues to be contested and elaborated, coming to be understood in terms of various aspirations, including "anti-speciesist, anti-racist, environmental, or health-centric" projects (Wrenn, 190). .

Analytically separating these three different contexts (i.e., identity articulation, identity conflict, and identity creation) is bound to seem somewhat artificial. In practice they tend to operate together and overlap in important ways: conflicts may lead to identity articulation; creation may generate new conflicts, and so on. My purpose in distinguishing between them is to draw attention to the fact that different deliberative contexts concerning practical identities require different kinds of argumentative strategies, problem-solving techniques, modes of creativity or inventiveness. These divergent strategies, I maintain, bear a kind of theoretical correspondence with moral individualism and humanism. Each kind of deliberative context yields a different way of thinking about the concept of moral status (at least—as I have argued—in so far as our practical identities are tethered to our ascriptions of moral status). Noting these differences allows one to elucidate different possible strategies for engaging in these kinds of moral reflection. These differences, I want to suggest, explain the functional diversity of concepts like that of moral status, and hence the motivations driving individualists and humanist in opposing directions.

How does acknowledging these diverse deliberative contexts yield theoretical correspondence? In short, I submit that (i) moral individualism can be understood as a strategy for dealing with distinctive kinds of identity conflicts, namely, “no-overlap” conflicts in which there is no common ground on the basis of which that conflict might be resolved; (ii) that moral humanism can be understood as a strategy which links moral problem-solving to identity articulation; and (iii) that whereas individualism reflects a thin-to-thick conception of identity creation, their humanist opponents go in for a thick-to-thin model.

Consider first some of the core features of moral individualism. As I described it in the previous chapter, its starting point is to deny the relevance of “category membership” to questions about what kinds of treatment various entities are owed, morally speaking (McMahan 2005, Rachels 2005). Motivating this denial, I argued, was a commitment to the idea that an agent’s reasons for attributing moral status must be in some sense, attitude-independent. This commitment, in turn, directs individualists to specify a set of non-moral properties or relations that seem intuitively relevant to questions of moral treatment—the idea being that these properties or relations provide an attitude-independent grounds for moral status claims. As Jeff McMahan explains, moral individualist’s are interested fundamentally in “status-conferring intrinsic propert[ies]”, which give “its possessor a moral status that is a source of ‘agent-neutral’ reasons – that is, reasons that potentially apply to anyone” (McMahan 2005, 355).

I want to suggest that one can regard moral individualism, at least in part, as recommending a deliberative strategy motivated by cases of identity conflict in which an intermediary practical identity is lacking (this could emerge as either an internal or external disagreement). Not only do these types of disagreements lead to practical conflicts about which courses of action are morally required; but, as I have suggested, they often take the form of disagreements about the scope of an agent’s obligations. In the latter type of case, practical identity conflicts just are disagreements about moral status wherein both parties (or a single person, if we are talking about conflict in the internal sense) are unable to appeal to a shared practical identity on the basis of which to resolve the disagreement. Moral individualism, I am suggesting, represents a strategy for resolving these types of practical conflicts. A strategy which involves appealing to a set of traits which are supposed to generate moral reasons independently of the practical identities at stake in the given conflict.

The individualist's deliberative strategy can, moreover, take either a strong or a weak form. The former attempts to resolve practical identity conflicts *via* an appeal to status-conferring properties which generate reasons independently of any practical identity to which one or more parties is antecedently committed. In its weaker form, the individualist's deliberative strategy essentially appeals to status-conferring properties which those experiencing the conflict are likely to regard as relevant, given some other practical identity to which they are committed.

Humanists, by contrast, emphasize that moral interpolation is often a matter of extending, and reinterpreting our practical identities. Rather than appeal to some set of neutral status-conferring properties as a means of resolving identity conflicts, these theorists hold that an existing commitment to a practical identity always allows for further reflection. In other words, Humanists are sensitive to the ways in which identity articulation serves as a means of problem-solving or moral edification. Consider, for example, the distinction between "contemplation" and "observation" to which Cora Diamond appeals in discussing how her own philosophical position provides reasons in support of moral vegetarianism. Whereas for individualists like Singer, such reasons are generated by a kind of "observation" (i.e., that certain biological capacities are morally salient), Diamond construes ethical persuasion as a kind of "contemplation" of the richness and the internal diversity of certain conceptual practices (and those centered on the notion of *human beings* in particular).

It is this view towards moral persuasion that motivates Diamond's appeal to poetry as a medium for spurring reflection about our attitudes towards non-human animals. For example, the point of Jane Legge's piece of "vegetarian propaganda" entitled, "Learning to be a Dutiful Carnivore" Diamond suggests, isn't to tell people how to behave or to dictate their feelings, rather, it addresses itself to people who already are disposed to have a complex range of background feelings and attitudes (Diamond 473). The poem encourages conceptual "contemplation" in at least two respects. First in leading the reader to reflect upon the tensions within her own commitments, sentiments, responses, and dispositions. The second—and Diamond is especially concerned with this—is to highlight our capacities (and incapacities) to extend elements of moral practice to new cases—to come to see animals as "companions" for example, or as the proper object of pity (478). The point is that this kind of moral contemplation, for Diamond, necessarily begins with a rich practical viewpoint and tries to develop it from within in order to bring about a kind of ethical transformation. Another way of putting this point is that the humanist's central concern is with the fact that our practical identities are always given to further articulation. There is an ever-present possibility that they can be worked on from the inside—especially as a means of self-transformation and moral persuasion.

Finally, moral individualism and humanism lend themselves to quite different models of practical identity creation. Moral individualism coheres with what I am calling a thin-to-thick model of practical identity creation. The individualist's primary concern is with determining antecedently, as it were, the obligations that a person's moral identity should be constructed around. Indeed, moral individualists are seldom (if ever) interested in questions about the broader social significance or narrative descriptions associated with a given practical identity. If they are, these considerations are understood as secondary and either subordinated to the primary obligations generated by the discernment of morally-relevant properties, or even problematic forms of sentimentalism which preclude moral reasoning.

Moral humanists, by contrast, can be read as emphasizing what I am calling a thick-to-thin model of practical identity creation. That is, humanists tend to emphasize the priority of practices to obligations, suggesting a model on which grasping an obligation is the result of reflection on some extant practical identity (as opposed to a model on which practical identities are generated by a grasp of some antecedent set of obligations). This follows from the humanist's claim that obligations cannot be understood independently of a conceptual understanding the kinds of things to whom they are supposed to be directed (e.g., an understanding of the human-directed obligations is, in part, constitutive of the concept *human being*). Given that such conceptual understanding presupposes on-going participation in a form of life, for humanists, therefore, obligations are best understood as dependent on a prior set of practices and conceptual understandings that make up a practical identity.

I have been arguing that the concept of moral status is closely connected to the notion of a practical identity, and that there are a multitude of deliberative contexts—i.e., elaboration, conflict, and creation—within which people work out their practical identities. These deliberative contexts not only constitute important types of moral interpolation, but they also generate a plurality of strategies for arriving at moral status attributions. Finally, I suggested that the availability of these strategies affords an explanation of some of the key difference between individualists and humanists. Whereas the former can be read as a response to the ever-present possibility that moral conflicts may lack a common ground, the latter can be read as a response to the ever-present possibility of expanding, or transforming one's practical identity. While what I have said about this explanatory strategy is admittedly just a sketch, my hope is that it throws some light on the problem of intractability. By attending to the connection between moral status claims and our evolving practical identities, one can see the disagreement between humanists and individualists from a new perspective, namely, one in which they reflect distinct strategies for mediating the interpolation of practical identity.

6.3.2 Moral Status and Moral Progress

In taking up the relationship between moral status and practical identity, I relied on the idea that there could arise a multitude of deliberative contexts which—at least if I am right—correspond (in some broad sense) with competing individualist and humanist approaches to moral status. That discussion can be thought of as a primary elaboration of the state-of-nature model, in so far as it did not turn on the content of any (historically conditioned) practical identity in particular.[43] By contrast, I now want to foreground, at least in some very cursory way, a secondary elaboration of the state-of-nature model by examining an important (yet, to my mind overlooked) set of conceptual connections between moral status and the notion of moral progress. What makes the following discussion a secondary elaboration is that it identifies a set of conceptual practices that have been elaborated historically. In particular, the conceptual connection between moral status and moral progress with which I shall be concerned, is largely endemic to the history of liberal societies.

[43] Although some of the examples I used suggest avenues for a more historically detailed secondary elaboration of the practical connection between moral status and moral identity. For example, a history of the development vegetarianism and veganism as distinct moral identities could yield insight into this practical connection.

Despite this important difference with the preceding discussion, the structure of my argument shall remain largely the same. Just as different forms of deliberation about practical identity call for divergent strategies for conceiving of moral status, so too, I shall argue, there is variability in terms of how the relationship between moral progress and moral status is currently understood. These differences, in turn, correspond to (and help to explain) the theoretical differences between individualists and humanists.

Following the editors of a recent special issue on the topic, the idea of moral progress can denote (at least) two distinct clusters of issues. The first involves a comparative evaluation of two states, and the second concerns a strategy for social change (Musschenga and Meynen 2017, 3). In what follows I am eager to examine how the concept of moral status has come to play a role in connection to both senses of moral progress.

Few would deny that "moral progress" is a notion that is subject to historical inflection. Even those who are skeptical of its existence would likely accept that at different times and places, people have taken "moral progress" to signify different things—whether it involves perfection along religious lines, the resolution of class conflict, or as a number of 19th and early 20th century Europeans seemed to think, an invidious process of increased "civilization" (Musschenga and Meynen 2017, 4).[44] While much could be learned from an historical investigation into these diverse conceptions of moral progress, I shall be concerned primarily with the much more limited task of sketching a conception of moral progress that has emerged within our own post-war human rights culture: the idea of an ever-expanding circle of moral concern. This is a conception adopted by many writers who—arguably, and quite strikingly—

[44] Indeed, one of the most obvious strategies for arguing against the very idea of genuine moral progress involves pointing to the variety of ways in which the concept has been understood.

take on quite distinct philosophical projects and commitments. As Michele Moody-Adams points out, writers as diverse as Peter Singer, Richard Rorty, and Martha Nussbaum adopt this conception of moral progress (Moody-Adams 2017, 154). Others take this conception of moral progress to be something like a common sense understanding of the notion.[45]

One reason for this exclusive focus is that this more local conception of progress is most clearly connected to the concept of moral status. Indeed, to expand the circle of moral concern just is—in large part—to attribute moral status to a greater number of beings. In taking up this local conception, I shall argue that it is far from univocal. The notion of an expanding circle of concern admits of various "theoretical choice-points" and can be mobilized for different purposes. It is within this functional variability that I aim to locate further explanatory resources for overcoming the problem of intractability.

As a number of writers have noted recently, the concept of moral progress is quite complex. First, one can understand the notion in terms of either *social* or *individual development*. The former view thinks of moral progress at either the level or populations, institutions, or societies, whereas the latter view looks at changes in moral attitudes, dispositions, behavior, or values that a person undergoes (Schinkel and de Ruyter 2017, 124-5). One might ask, for example, whether a government has made moral progress by looking at its policies over a period of time. Alternatively, one might ask whether an individual has made moral progress over a period of their life. Both senses of progress can—though need not—be framed in terms of the expanding-circle model. The government might be said to have made moral progress in so far as

[45] Allen Buchanan and Russell Powell take this dimension of moral progress as the focus of their influential "naturalistic theory of moral progress" (2016). Their description of this idea invokes explicitly the notion of moral status. As they put it, "We will focus on one dimension of moral progress, namely, the movement toward increasingly 'inclusive' moralities, or expansions of the sphere of beings that are regarded as having moral standing or equal basic moral status" (985).

it has recognized the moral or legal status of a broader range of people. The individual might be said, similarly, to have progressed morally by expanding the scope of those to whom they take themselves to have obligations.

Second, Anders Schinkel and Doret de Ruyter draw a helpful distinction between *weak* and *strong* conceptions of moral progress. Whereas the former denotes some positively evaluated change, the latter involves the "the external expression of an internal or underlying change" which allows the change to be regarded as stable and non-superficial (Schinkel and de Ruyter 2017, 123). In part, one's ability to discern instances of weak from strong progress requires knowledge of the background conditions under which it occurred. For example, one might be less inclined to view as strong moral progress a corporation's implementing a set of progressive policies simply because their competitors are doing the same, than one which adopted those policies as the result of careful collective deliberation. Likewise, a person who becomes kinder or more generous because their team won the big game could be less plausibly said to have made progress in the strong sense than could the person who has become kinder and more generous as the result of having undergone some change in sentiments through, say, reading Dickens novels. A key component, according to these authors, is that the strong sense of progress involves a certain degree of irreversibility (122).

Third, one might distinguish between *teleological* and *problem-based* conceptions of progress (Schinkel and de Ruyter 2017; Kitcher 2017).[46] On the former, progress consists in a change that is evaluated in relation to some fixed goal or end-state, which can be specified fairly

[46] Schinkel and de Ruyter contrast, following Godlovich, a teleological conception of progress with a notion of "improvement" (123). A closely related distinction is between formal or substantive criteria for evaluating moral progress.

clearly in advance (Schinkel and de Ruyter 2017, 123-4).[47] A problem-based conception allows for a more malleable set of criteria according to which change is measured, and need not presuppose a clear picture of what the finished state of some process might look like. Kitcher puts the contrast in terms of "progress to" versus "progress from" (Kitcher 2017, 48). A simple case of (non-moral) teleological "progress to" would be "travelling to an intended destination" (48). Working out defects in existing technologies, or finding new treatments for diseases are paradigms of a more pragmatic, or problem-based "progress from."

Having outlined some of the internal diversity within the idea of moral progress as a circle of expanding concern, I now want to argue that moral individualists and humanists tend to navigate these different theoretical choice-points in contrasting ways, resulting in diverging ways of thinking about progress and its relation to moral status.[48]

Consider first, the question of whether to think of progress at the level of institutions or as a question of individual development. Moral individualists tend to be concerned primarily with the question of moral progress on the institutional level. This commitment is illustrated in their concern with speciesism as a kind of moral discrimination affecting social attitudes. Although there is a sense in which individuals can be guilty of speciesism, those who employ the term (e.g., Singer, McMahan, and Regan), tend to present their critiques as directed at population-level beliefs, social practices (e.g., factory farming, vivisection), or, more generally

[47] Or as Kitcher puts it, "Teleological concepts of progress do posit a goal, and take progress to consist in diminishing distance from the goal" (48). By contrast, "Pragmatic progress… lies in solving problems" (49).

[48] In emphasizing these three sets of differences my aim is descriptive—rather than specify whether, for example, a teleological view of progress is superior to a problem-based one, I am inclined to insist that these differences simply alert us to the fact that our conception of progress is quite complex. Of course, we can debate the merits of a particular way of thinking about a particular issue—but on my view, there is little to be gained by arguing which uses of the concept of moral progress are foundational or have some kind of priority independently of their application.

“official morality” (Singer 2009, 571). The individualist’s concern lies not so much in what it would mean for any given person to make progress in terms of their moral development (although the individualist will think of this process in terms of increased rationality or consistency in beliefs), but in terms of an expanded circle of moral concern at the level of societal values that are embodied in institutions and social practices. When these writers do take up questions about individual development, they do not address particular people, but rather “agents” whose particular level of moral development and background commitments are not taken as relevant. Although some moral individualists do admit that a person’s “special relationships” with others can provide a source of normative reasons, they tend to treat such reasons as only instrumentally valuable, thereby subordinating them to a kind of cost-benefit analysis undertaken from an agent-neutral perspective. For example, while McMahan acknowledges that parents may come to have relation-dependent obligations towards their severely disabled child, he thinks that the value of these obligations needs ultimately to be weighed against a set of broader social considerations. As he puts it, “If we compare the number of radically cognitively impaired human beings who benefit from our partiality with the number of animals who suffer from our tendency to regard them primarily as means to our ends, it is hard to believe that the effects of species partiality are desirable overall from an impartial point of view” (McMahan 2005, 361).

By contrast, moral humanism lends itself better to an account of moral progress which targets individual development. The humanist’s focus is not on identifying a set of properties which will determine in advance how the circle of concern ought to be expanded *from some disinterested perspective*. Rather, their focus is on something like moral edification arrived at through cultivating a richer understanding of one’s concepts and practices. This is fundamentally a reflective process which cannot be undertaking by institutions or populations.

These differences in focus lead individualists and humanists to different answers about whether a weak or strong account of progress should be given priority. Again, this view mirrors my approach above.

Due to their institutional focus, individualists tend to adopt a weak or thin notion of moral progress (i.e., change plus normative judgments). This is evinced by their tendency to downplay the background conditions according to which an agent’s moral development might be understood. The moral individualist’s aim of offering a theory of the grounds of moral status is to specify a set of non-moral properties which determine how far the circle of concern is to be expanded. As I argued in the previous chapter, individualists typically see as irrelevant (or secondary) considerations involving *how,* psychologically*,* this expansion is to be realized.

A focus on progress-as-individual-development leads humanists to adopt a strong notion of moral progress. This means that they are not simply interested in a means of evaluating change, but also in background features of a situation, according to which the change can be seen an externalisation of some internal transformation. As I explained in Chapter Five, humanism involves both sensitivity to context and to a person’s affective and imaginative capacities.

Finally, whereas moral individualism is best thought of as invoking a teleological conception of progress, humanists tend to operate with a more localized, problem-based conception. On the one hand, moral individualism can be understood, in part, as an attempt to generate a set of criteria that can determine whether the circle of concern has been expanded in the right way. A central function that any theory of the grounds of moral status is meant to serve, is to specify a set of non-moral properties which will enable one to determine which kinds of things are the proper objects of moral concern. Progress occurs when the limits of concern align with the dictates of the theory.

By contrast, insofar as they are committed to the project of expanding the circle of moral concern, moral humanists do not seem to be interested in providing a stable set of criteria that will enable one to tell in advance whether this expansion has been carried out far enough. This is because they tend to invoke an improvement or problem-based conception of progress which acknowledges the complexities of moral conflict, and which attempts to work through those conflicts by using the resources available within a more narrowly circumscribed context. This is, one might think, reflected in the humanist's attention to the kind of progress we observe in child-rearing.

6.4 Conclusion

In this chapter, I have developed an expressivist account of moral status which proceeds *via* the methodological guiderails of a pragmatic genealogy. My central argument has been that adopting this perspective can help resolve the problems of intractability and eliminativism, facing current accounts of moral status. Beginning with an abstract state-of-nature model, I suggested that the concept of moral status answers to a set of generic needs resulting from the fact that our moral practices involve a phenomenon which I called interpolation. I then showed how the concept has become integrated into other deliberative contexts—beginning with more rudimentary ones in which agents disagree about the scope of their obligations or need to make moral generalization, before turning to how the concept of moral status could be expected to get taken up within deliberative contexts concerning practical identity and moral progress. This strategy, I claim, offers a piecemeal response to the eliminativist's worry that *moral status* is a dispensable notion.

Similarly, my solution to the problem of intractability has been holistic. The more attention paid to the potential uses to which the concept of moral status may be put, the less reason for thinking there is a univocal concept which must be "grounded" in a set of properties (*pace* individualism), or understood exhaustively through reflection on a single concept such as "human being" or "person" (*pace* humanists). I have also argued that the pragmatic genealogy offered here can retrodict features that our current concept of moral status displays. For example, I have appealed to the concept's potential connection to a broad range of practical concerns to make sense of the diversity internal to moral individualism, and I have suggested that some of the central differences between individualists and humanists result from their attunement to different ways that the concept of moral status gets connected to questions of practical identity and moral progress.

The argument deployed in this chapter suggests that the concept of moral status is an inextricable feature of our moral lives. At the same time, I do not take this account to have vindicated fully our conceptual practices. Drawing attention to the concept's functionality within a state-of-nature model can help explain why it would have emerged. And drawing attention to the functional diversity that it could be expected to have taken on can help explain some of the theoretical disagreement within contemporary ethics. But further work would need to be done to lend robust normative support for any given conception of moral status that one might adopt. For example, my effort to elucidate the different contexts in which thinking about practical identity yields a conception of moral status that explains the existence of those strategies, but does comparatively little to make sense of their value. Addressing such normative questions requires that we, at least in part, turn to history. I take this up, albeit by way of example rather than through a full-blown normative framework, in the next chapter.

CHAPTER 7

A HYBRID ACCOUNT OF ROBOT RIGHTS

7.1 Introduction

I have argued that our conceptual practices of attributing moral status serve several functions. On the one hand, they enable us to regulate our moral economy and to engage in moral interpolation. On the other hand, they answer to more historically local practical demands: first, by helping us deliberate about our practical identities; and second, by enabling us to articulate a conception of moral progress centered on the idea of expanding the circle of moral concern.

So far, however, I have said little about this pragmatic genealogy's normative upshots. Can an inquiry into to the function and etiology of our conceptual practices do more than simply describe how we, in fact, attribute moral status? Can it correct moral practice by, for instance, telling us how we *should* make such attributions? Or is a pragmatic genealogy the sort of philosophical approach which—to borrow a phrase from Wittgenstein—necessarily leaves everything as it is? (Wittgenstein 1953/2009, 55).

There are good reasons to desire normative guidance from a pragmatic genealogy. Consider this issue in the context of contemporary technological developments. Advances in AI and engineering may soon revolutionize the boundaries of our moral communities. The effects of integrating social robots into our homes, workplaces, schools, hospitals and labs, transportation sector, military, and entertainment industries will almost certainly be tremendous.[1] The more sophisticated these machines become—as they increasingly display intelligence, emotions, autonomy, creativity, and sapience—the deeper and more complex the social relationships we are likely to form with them. Some philosophers have no trouble envisioning near-future scenarios in which we have extended moral consideration to intelligent machines (Bostrom 2014; Floridi 2002; Gunkel 2014; Tavani 2018). On this view, robots acquire human-like capacities we shall owe them certain forms of respect, and find ourselves weighing their interests against our own. Future engineers may come to have the kinds of obligations towards their robotic inventions that parents have towards their children (Schwitzgebel and Garza 2015; 108-9).[2] Other philosophers doubt that these scenarios are likely to occur—at least anytime soon (Andreotta 2021; Mosakas 2021; Müller 2021). For them, robots are artifacts. No matter how complex their behavior becomes, our obligations to them will never differ significantly from those we have to our toasters. Of course, owing to our own psychological tendencies to project mentality or agency onto social robots, we may risk mistakenly ascribing moral status to them. But these are outcomes to be avoided, perhaps by implementing design principles which limit the degree to which machines are intended to physically resemble humans (Bryson 2009).

These nascent philosophical disagreements have already entered public discourse. In 2017, an interactive social robot named Sophia became the first entity of its kind to be granted

[1] For a general discussion of the potential impact of robots within our lives, see Bostrom (2014), Carr (2014), Darling (2014; 2021), Nørskov (2016). For a discussion of the economic impact of integrating robots into the workplace, see Ford (2015), Danaher (2017). Some writers have considered features of human-robot relations, especially sexual and romantic relations with robots (Danaher 2019; Frank and Nyholm 2017), but also friendship (Danaher forthcoming; Marti 2010; Darling 2017) and care-giving (Sharkey and Sharkey 2010). Since the European Parliament's Committee on Legal Affairs issued a 2017 report proposing the creation of the category of "electronic personhood," there has been considerable discussion of the legal status of social robots. For an overview of this debate, see Parviainen and Coeckelbergh (2021).

[2] In addition to the parent-child relationship, other relationships that have been used to analogize the human-robot relation include the employee-employer relation, or the god-creature relation.

citizenship. This decision—by the Saudi Arabian government—provoked a range of reactions from celebration to condemnation (Parviainen and Coeckelbergh 2021). These controversies suggest that we shall soon face a moral reckoning concerning the place that SRs occupy within our moral communities (Danaher 2019).[3]

In this chapter, I argue that pragmatic genealogy can provide normative guidance in these debates. My claim is that the model developed in the previous chapter vindicates a *hybrid approach* to attributing moral status to social robots. Before describing this position, allow me to situate it within a range of possible normative stances on moral status ascription. These stances are defined relative to the two central debates that I outlined in Chapter Five.[4]

On the one hand, there is a range of possible *first-order* commitments concerning the properties that ground moral status claims. Many theorists hope to secure normative guidance through ontological inquiry into status-conferring properties that generate normative reasons. First-order disagreement about moral status concerns *which* properties are, in fact, relevant.

On the other hand, there are competing *second-order* or metatheoretical views involving the deliberative strategies agents ought to adopt when ascribing moral status. I have used the terms *moral individualism* and *moral humanism* as labels for the two sides of this debate. Moral individualists contend that whether an entity has moral status is a function of its properties. In other words, these writers argue that answers to the first-order questions are necessary for ascribing moral status. Moral humanists, by contrast, deny that inquiry into the first-order

[3] Danaher suggests that the rise of social robots may also produce a "crisis of moral patiency" which would "compromise both the *ability* and *willingness* of humans to act in the world as responsible moral agents, and consequently could reduce them to moral patients" (Danaher 2019, 2).

[4] Insofar as eliminativism about moral status is a live option, the present account has already made some normative headway given that it tells us that *some* conception of moral status answers to very general practical needs that we have.

questions is a fruitful stating point. They seek to prioritize other forms of reflection when it comes to deliberating about the scope of one's obligations.

Across these two debates, one can draw the following distinctions:

> *First-order monism*: there is a unique set of intrinsic properties which ground moral status claims.
>
> *First-order pluralism*: there is a multitude of (potentially competing) intrinsic properties which ground moral status claims.
>
> *Second-order monism*: moral individualism and moral humanism are mutually incompatible. One can and should only accept one, but not both.
>
> *Second-order pluralism*: one can and should accept core tenets of both individualism and humanism.

In principle, a pragmatic genealogy could seek to vindicate any of these normative stances. For example, it might aim to establish first-order monism by advancing substantive views about which properties are, in face, status-conferring.[5] Alternatively, it might aspire to a version of second-order monism by vindicating a particular deliberative strategy (i.e., individualism or humanism) for reasoning about moral status attribution. In this chapter, I shall argue that a pragmatic genealogy of moral status supports pluralism on both fronts. That is, the functional and etiological hypotheses defended in the previous chapter vindicate pluralism with respect to both the first and second-order debates. I call this position a *hybrid account of moral status ascription*.

[5] For example, a pragmatic genealogy might posit that the concept of moral status performed function F, and that it was only by ascribing moral status on the basis of property P that F could be fulfilled. Given what I have argued in the previous chapter, such a view seems highly implausible—my point is just to indicate a possible theoretical orientation. Someone might object, however, that insofar as a pragmatic genealogy aspires to indicate the functional plurality of our concepts, it is therefore incompatible with *any* monist account of normativity. While this may be true in many (perhaps most) cases, it is important to recognize that it is, in principle, possible for a pragmatic genealogy to reveal functional unity within our conceptual practices. It is a contingent feature of our concepts that they tend to have a variety of uses and purposes. This is something that pragmatic genealogies are good at elucidating—but it is not a necessary feature of the method that it will always yield a pluralist perspective.

Given that monism has been the default position within applied ethics, this hybrid view turns out to be a radical proposal. Indeed, a likely objection to this view stems from the apparent mutual exclusivity between individualism and humanism. That is to say, the default view is that adopting a humanist standpoint requires rejecting the individualist's insight that there are status-conferring properties, and *vice versa*. In response, I argue for the compatibility of these two positions by expanding on a central result of the previous chapter—namely, that moral individualism and moral humanism track different deliberative strategies for attributing moral status. Construing the two positions as answering to separate, but individually important aspects of our practices, allows them to be understood as complimentary. So far, however, I have left it an open question *which of the strategies ought to be endorsed or adopted?*[6] I submit that the acceptability of both second-order positions depends on whether the deliberative strategies they involve are worth cultivating, continuing, and endorsing. And one's answer to these questions, I suggest, following Matthieu Queloz, depends on one's conception of an agent (Queloz 2021, 236-42). In other words, the acceptability of humanism or individualism turns on the question of what kind of agents we are or aspire to be.

The hypothesis explored in this chapter is that we are increasingly subject to a phenomenon that I call *practical identity proliferation*. As members of globally-connected liberal democracies, we are the kinds of agents who must increasingly navigate between a host of practical identities and who must contend with the frequent and pronounced identity conflicts accompanying them.[7] Identity proliferation gives rise to new forms of interpersonal

[6] In other words, so far I have argued that there is space for individualism and humanism to coexist within moral practice, but I have not argued that there are good reasons to adopt either of the two metatheoretical stances and their associated deliberative strategies.

[7] For a discussion and defense of a cosmopolitan ethos, see Appiah (2006).

disagreement and generates novel challenges when designing public institutions. My claim is that a hybrid approach to moral status ascription—encompassing both the first and second-order pluralism described above—is the most reasonable response to these practical problems.

Section 7.2 explores how moral individualism and moral humanism have animated recent debates about the moral status of social robots, and argues that both approaches are limited by: (i) an underlying commitment to monism, and (ii) a failure to recognize the connection between moral status attribution and practical identity formation. In particular, I argue that moral individualism ultimately fails to establish that moral properties can provide reasons for action that apply universally, or to any and all moral agents (*the problem of relevance*). I also contend that moral humanism ultimately fails to account for the fact that appeals to such properties *do* often factor into moral justification—albeit in a manner that is defeasible and much more contextually determined than individualists tend to think. In Section 7.3, I introduce the historically local problem of practical identity proliferation and argue that it motivates a hybrid approach to moral status attribution. Given that we are the kinds of people who must navigate a world of proliferating, evolving, and conflicting identities, we require a multitude of deliberative strategies for ascribing moral status. Finally, I present a set of positive proposals for how to conceive of the moral status of social robots along hybrid lines. I argue that moral individualism presents a set of deliberative strategies involving *value simplification* that are extremely useful within public institutions. By contrast, moral humanism presents a set of deliberative strategies that preserve the complexity of our values. These complexity-preserving strategies are crucial for individual moral development and for facilitating moral transformations within communities.

7.2 The Moral Status of Social Robots

Following Kate Darling, a social robot is "a physically embodied, autonomous agent that communicates and interacts with humans on a social level (Darling 2014, 2).[8] This definition is meant to exclude inanimate computers (e.g., software or search engines), as well as robotic machines that are not designed to interact with humans.

Recent debates about the moral status of social robots (SRs) are attractive sites for fleshing out my argument for a hybrid approach for at least two reasons. First, they clearly encapsulate both the first-order and second-order debates with which I am presently concerned. That is, not only do scholars disagree about which properties ground moral status claims (first-order concern), but there is widespread metatheoretical disagreement—reflecting the divide between moral individualism and humanism—about how moral inquiry ought to proceed (a second-order concern). Second, it seems plausible that advances in natural language processing will enable social robots to participate in an ever-widening array of socio-linguistic practices, and ultimately, to share in practical identities with us. As they become active participants in shaping our shared moral lives, it is plausible to think that social robots will begin to demand rights of their own (Brooks 2000). This suggests an important respect in which social robots differ—at least in significant degree—from non-human animals. Many writers argue that while animals have moral status because they are *sentient*, human beings enjoy a higher degree of moral status due to their possession of some additional capacity—*sapience* (Bostrom and

[8] Darling posits three additional features that distinguish social robots from other objects such as toasters. In particular, these features explain why we tend to project psychological traits onto social robot, and why they tend to elicit emotional responses from us. First, the fact that social robots are *physically embodied* suggests that we tend to interact with them in ways that we would not interact with a virtual object on a screen. Second, our tendency to attribute mentality to social robots likely stems from the fact that their behavior often appears autonomous or self-directing. Finally, social robots exhibit social behavior in the sense that "they are also able to mimic cues that we automatically, even subconsciously associate with certain states of mind or feelings" (Darling 2014, 6).

Yudkowsky 2014, 322-2). Social robots complicate this picture. On the one hand, it seems likely that they will someday possess sapience—or at least pass any behavioral test for sapience we can devise—suggesting that they would possess a higher degree of moral status than non-human animals. On the other hand, there are grounds for doubting that social robots will become sentient, which suggests the opposite conclusion.

7.2.1 Moral Individualism: For and Against the Moral Status of Social Robots

Most philosophers concerned with the moral status of social robots subscribe to moral individualism—or as it is often called—a property-based view (Andreotta 2021, 24; Gordon 2021, 463; Mosakas 2021, 430).[9] According to this position, whether a social robot has moral status depends on whether it possesses of morally relevant intrinsic properties, such as sentience, sapience, intelligence, rationality, among others.

Moral individualism invites (at least) three sets of questions: *ontological questions* about which properties are status conferring, *empirical questions* about whether social robots can (or do) instantiate these properties, and *epistemological questions* about whether we can know that SRs instantiate them. If it turns out that these questions currently lack definitive answers, moral individualism is consistent with arguments both for and against the claim that social robots may someday possess moral status.

Consider the view—call it *robot rights optimism*—that social robots either currently, or will likely someday possess moral status. Eric Schwitzgebel and Mara Garza advance an

[9] Even those who reject moral individualism acknowledge that it is the standard approach for dealing with the moral status of SRs. For a critical discussion of this claim, see Coeckelbergh (2010; 2018, 147), Gunkel (2014).

argument along these lines, called the *No-Relevant-Difference Argument* (NRDA) (Schwitzgebel and Garza 2015, 99):

> P1 If A deserves some degree of moral status and B does not deserve the same degree of moral status, then there must be some relevant difference between A and B that grounds the difference in moral status.
>
> P2 There are possible AIs or social robots who do not differ in such relevant respects from entities (e.g., human beings) who currently possess moral status.
>
> C1 Therefore, there are possible AIs who deserve some degree of moral status. [10]

The first premise of the NRDA is entailed by moral individualism. Schwitzgebel and Garza do not offer an extended defense of it, but simply insist that its denial would render "ethics implausibly arbitrary" (Schwitzgebel and Garza 2015, 99). I shall return to this point below.

The second premise is general enough to permit disagreement about which properties ground moral status and to allow agnosticism about whether social robots will someday display those properties (given, for instance, reasonable expectations about technological innovation).[11] Perhaps unsurprisingly, optimism that social robots either presently, or will soon possess moral status tends to be a function of the properties one takes to be status conferring.

Schwitzgebel and Garza adopt a permissive "psycho-social view of moral status" according to which only psychological properties and social properties are status conferring (Schwitzgebel and Garza 2015, 100-1). They also defend the second premise against possible objections, noting that the claim is quite modest given its modality. While *some* artificially

[10] Coeckelbergh suggests an argument along these lines when we asks, "if (in the future) it turns out that robots share features with humans such as rationality or consciousness, then if we hold these features as a basis for human rights, why restrict those rights to humans?" (2010, 211). More generally, the NRDA can be taken as a version of the argument from species difference (or, as it is often called, the argument from marginal cases).

[11] Some writers take the issue to be whether social robots will *ever* possess status-conferring properties, whereas others focus more on the question of whether social robots will likely soon possess those properties.

intelligent entities (e.g., smartphones) appear not to instantiate morally relevant psycho-social properties, it is difficult to argue that *no* artificially intelligent entity will ever (or will likely soon) instantiate such properties. As these authors observe,

> no *general* argument has been offered against the moral status of all possible artificial entities. AI research might proceed very differently in the future, including perhaps artificially grown biological or semi-biological systems, chaotic systems, evolved systems, artificial brains, and systems that more effectively exploit quantum superposition (104).[12]

While Schwitzgebel and Garza articulate a relatively standard individualist position within debates about the moral status of SRs, others argue for an even more permissive view. Luciano Floridi has advanced a stronger—and much more controversial—framework, *information ethics*, which affords some degree of moral status to anything that can be considered an information object (Floridi 1999; 2002).[13]

Other moral individualists—call them *robot rights skeptics*—deny that social robots will likely soon (or ever) possess status-conferring properties. These writers accept the first premise of the NRDA while denying the second. In its strongest form, this kind of skepticism amounts to the claim that it is metaphysically impossible for social robots to possess moral status.[14] One might argue for such a view on Aristotelian grounds. Perhaps, each kind of thing is defined by its

[12] For a discussion of the likelihood that these technologies will be available in the (relatively) near future, see Bostrom (2014).

[13] For critical discussions of Floridi's information ethics—especially as it bears on the questions of robot rights—see Brey (2008), Coeckelbergh (2010, 217), Gunkel (2014, 122-6) and Mosakas (2021, 436-8).

[14] There is something rather paradoxical about the attitudes associated with this strongest form of skepticism. On the one hand, many writers suggest that it represents a default or common-sense stance towards the question of robot rights. Most people, that is, find granting moral status to artificial entities to be an utterly absurd idea. On the other hand, as Schwitzgebel and Garza's remarks cited above indicate, it is difficult to find examples of philosophers who actually subscribe to or defend this strong of a position (Schwitzgebel and Garza 2015, 101). It may be the case that our *present* technological capacities fall short of actualizing artificial consciousness—but why should we think that such interventions would be impossible in some strong metaphysical sense?

telos; such that it is metaphysically impossible for an artifact—however sophisticated—to have the kind of telos that would suffice for its possessing moral status.[15] In Book 7 of *Metaphysics*, Aristotle distinguishes between natural and artificial entities, suggesting that whereas the former have their origin "in themselves" the latter are always dependent on the intentions of a craftsperson—hence, on human intervention (1032a-1035a).

Setting aside any commitment to supernaturalism that teleological accounts risk incurring, there are several problems with such a position. First, as others have argued, any *a priori* distinction between *artifactual* and *natural* entities is of scant use to the natural and social sciences (Miller 1994). Hence, the view runs afoul of even a weak form of methodological naturalism. Second, even if one were to grant that things are defined by their ends or purposes, these appeals only justify moral status ascription because they track other properties—such as practical reasoning or sentience—which end up doing the normative heavy-lifting (Schwitzgebel and Garza 2015, 101). For example, Aristotle's claim that the human *telos* as a life lived in accordance with reason and virtue is only a plausible basis for moral status because such a life involves the exercise of status-conferring properties that could, in principle, be shared by other entities.

A more plausible form of robot rights skepticism is engendered by what Adam Andreotta calls *the hard problem of AI rights*. Like many other writers, Andreotta takes phenomenal consciousness to be a necessary condition for moral status (Andreotta 2021, 24).[16] Our moral

[15] This claim could be understood in two ways. First, it might be taken to mean that an entity's *telos* is what determines its moral status, and that since no artificial entity could have the same *telos* as that of human beings, there could never be a case in which a non-human had the same level of moral status as humans. Second, someone might accept that some psycho-social property grounds moral status, while insisting that there is something about an artificially intelligent machine's *telos* that prevents it from ever realizing that psycho-social property.

[16] See also Mosakas (2021, 431).

concern for humans and non-human animals, for example, is justified only if such creatures undergo subjective experiences and are capable of suffering (24). Despite its popularity, this view seems to run up against a serious problem: given the absence of an agreed upon theory of what consciousness *is* or how it arises, how can one justify attributing moral status to *anything*? In other words, it seems difficult to even apply the consciousness criterion of moral status barring a solution to what philosophers of mind call *the hard problem of consciousness* (i.e., the problem of explaining why physical processes generate qualia).[17] When it comes to the moral status of other humans and non-human animals, Andreotta thinks there is a way around this issue. One can attribute conscious experiences to non-human animals for the same reason one can attribute them to one's *conspecifics*: because of behavioral, evolutionary, and biological similarities (Andreotta 2021, 25). But since there are comparatively few (if any) relevant evolutionary or biological similarities between humans and social robots, we cannot appeal to this explanatory strategy. As Andreotta observes,

> Given that advanced AIs will likely be constituted in ways that are very different to us… current approaches to animal consciousness do not map well to questions of AI consciousness. The 'Hard Problem' for AI rights… stems from the fact that we still lack a solution to the 'Hard Problem' of consciousness (Andreotta 2021, 19).

Although the hard problem of robot rights reflects a laudable naturalistic spirit that is absent from the Aristotelean argument, one can generate thought experiments which challenge the idea that evolutionary and biological similarities provide our only justification for attributing phenomenal consciousness to other entities (Schwitzgebel and Garza 2015, 103). Consider a phenomenally conscious human whose brain is gradually replaced by silicone chips. If, as seems intuitively plausible, minor replacement would not result in the loss of phenomenal

[17] For a discussion, see Chalmers (1995).

consciousness, then one should not expect a different result were the process to be carried out iteratively until the brain was transformed completely.[18] So long as the patient's behavior remained largely unchanged—and especially if they continued reporting the presence of conscious experiences—there does not seem to be grounds to deny that they would still be conscious.

In addition to these thought experiments, there are other ways for moral individualists to avoid the skeptical conclusion that the hard problem of AI rights entails. On the one hand, one might circumvent the biological-evolutionary explanatory strategy by appealing to empirical tests for consciousness. Recently, for example, Susan Schneider proposes several frameworks for testing whether a machine is conscious (Schneider 2019, chapter four).[19] On the other hand, an individualist could deny Andreotta's claim that consciousness is a necessary condition for moral status.[20] As mentioned above, Schwitzgebel and Garza's psycho-social view leaves open the

[18] This kind of thought experiment is often mobilized in support of the multiple realizability thesis—the claim that mental states can, in principle, be implemented on a wide range of physical bases. Andreotta does not find this line of argument convincing given its reliance on intuitions about cases for which we have no empirical support (Andreotta 2021, 27-8).

[19] The first, "AI Consciousness Test" is meant to serve as a sufficient, but not necessary condition for determining consciousness (Schneider 2019, 50). It attempts to "challenge an AI with a series of increasingly demanding natural language interactions to see how readily it can grasp and use concepts based on the internal experiences we associate with consciousness" (51). For critical discussions of Schneider's tests, see Andreotta (2021).

[20] John Danaher advances a view called *ethical behaviourism* (EB), which proposes the following sufficient condition for ascribing moral status to social robots: an entity has moral status "if they are roughly performatively equivalent to other entities that are commonly agreed to have significant moral status" (Danaher 1). So long as a robot "consistently behaves like another entity to whom we afford moral status, then it should be granted the same moral status" (5). If, for example, one is prepared to grant moral status to a dog on the grounds that he experiences pain, then according to the behaviourist, one ought to grant moral status to any entity that exemplifies pain behavior in (roughly) the same way that a dog would. On this view, "performative artifice" is sufficient. The question of whether there is something "going on 'on the inside'" is irrelevant, ethically speaking (3).

In contrast to moral individualism, which foregrounds the metaphysical question about which intrinsic properties are status-conferring, ethical behaviourism (like methodological behaviourism in psychology) is a "normative and meta-empirical thesis" (Danaher forthcoming, 9) concerning the kind of empirical evidence that should inform our theories of moral status. As Danaher notes, EB need not entail that metaphysical inquiry into the grounds of moral status is useful. The behaviourist can readily admit that, for example, sentience is the ontological basis for moral status ascription. Their point is that we ought to give priority to questions about our epistemic access possibility that social properties may be sufficient to ground the moral status of entities that lack conscious experience. And Floridi's information ethics allows that an entity can have moral status without being consciously aware (Floridi 2002).[21]

Before discussing objections to moral individualism, allow me to briefly mention two additional positions in the SR rights debate which proceed along individualist lines. Whereas the differences between the optimistic and skeptical views considered so far stem from metaphysical and epistemological concerns, these positions stem from the practical consequences of attributing moral status to social robots.

Joanna Bryson argues that, even if it were possible to design and build robots who have status conferring properties, the disadvantages of doing so would ultimately outweigh the benefits. For Bryson, we ought to regard and treat robots as our slaves (Bryson 2009). Not only would it be morally wrong to design robots to whom we owe moral obligations, but, according to Bryson, it would be wrong to allow people to treat robots as persons.[22]

Her main argument for this position is a consequentialist one: given that "humans have only a finite amount of time and attention for forming social relationships" (Bryson 2009, 5), we risk wasting this "precious commodity" by expending it on the kinds of "non-productive faux-social" relations with machines (5). More generally, misidentifying social robots as proper targets of moral concern incurs "economic and human consequences of time, money and

to those status-conferring properties—whatever they happen to be—and to claim that "inferences from behavior are the primary and most important source of knowledge about the moral status of others" (8).

[21] Neely argues that it is possible for AI to have interests even if they lack phenomenal consciousness, and that this suffices for their having moral status (Neely 2014).

[22] The obvious implication is that roboticists ought to avoid designing robots that mimic (or genuinely experience) human emotions or sentience.

possibly other finite resources being given to a robot that would otherwise have been spent directly on humans and human interaction" (6).

While Bryson may be right to caution against mistakenly allocating moral resources to non-sentient machines (given their apparent lack of moral status), her argument that we ought to refrain from designing robots who have moral status is unconvincing. First, even if this would result in fewer resources being allocated to humans, it is question-begging to assume that human interests would outweigh those of social robots.[23] Second, Bryson overlooks the potential advantages of human-robot interactions that would likely require anthropomorphizing design principles (e.g., where robots require some degree of human-like traits to perform educational or caregiving roles). At the very least, it is unclear what a consequentialist cost-benefit analysis would yield in these contexts.

Finally, several writers (including those who are skeptical of affording rights to SRs on individualist grounds) adopt *indirect* views of moral status ascription. On this picture, although we do not have moral obligations *towards* social robots, we have moral obligations *involving* them. In particular, we ought to refrain from mistreating social robots in order to avoid incurring harms towards ourselves and other moral agents (Darling 2014; Coeckelbergh 2018, 145; Cappuccio et al. 2019).[24] These positions can be considered forms of individualism because the very idea of an indirect duty depends on a contrast with direct duties—which are almost always understood along individualist lines. That is, indirect views typically begin by denying that social

[23] Andreotta raises a similar concern (Andreotta 2021, 22).

[24] In general, indirect views tend to come in one of two forms. Kantian versions of this view argue that we ought to grant social robots protection because doing so "may reinforce behavior in ourselves that we generally regard as morally correct, or at least behavior that makes our cohabitation more agreeable. It may also prevent desensitization towards actual living creatures and protect the empathy we have for each other" (Darling 2014, 19). Virtue-based versions of the argument point to the ways in which certain modes of interacting with SRs are more likely to promote virtue in humans (Coeckelbergh 2018, 145; Cappuccio et al. 2019).

robots do (or can) have moral status because they lack the requisite status-conferring properties.[25]

7.2.2 Problems with Moral Individualism

Despite its predominance within debates about the moral status of social robots, moral individualism faces several challenges. In this section, I critically examine these objections and argue that, while some are not as devastating as critics of moral individualism seem to think, others do present serious obstacles for a viable moral individualism about robot rights. I contend that the thrust of these criticisms suggests the need for a *constrained individualism*, which embraces a limited role for status-conferring properties within moral practice while rejecting the claim that these properties produce agent-neutral normative reasons. This position will serve as a premise in my argument for a hybrid approach in Section 7.3.

Mark Coeckelbergh argues that any version of moral individualism is bound to encounter "problems of *application*" (Coeckelbergh 2010, 212). Not only is it unclear what constitutes "respecting an entity's rights or capacity for suffering" (212), but any attempt to specify the ontological grounds of moral status is vulnerable to versions of the argument from marginal cases. That is, any such specification risks violating commonly held moral intuitions by either excluding certain entities who *appear* deserving of moral consideration, or by including entities that do not.[26]

[25] One exception to this is Coeckelbergh (2018), whose view I consider in detail below.

[26] I doubt that most moral individualists will find this point persuasive, given that they are free to 'bite the bullet' and deny, for example, that non-sentient human beings would have moral status. At times, Coeckelbergh runs dangerously close to begging the question against the individualists he criticizes. For example, he claims that, "Today robots are neither conscious nor sentient. It is even questionable if any of them really are (artificially) intelligent. This renders arguments based upon such features irrelevant to the problem of how to think about giving moral consideration to currently existing intelligent robots" (Coeckelbergh 2010, 212). It is not clear why an

Moral individualism also encounters *conceptual* and *epistemological problems*, which make it difficult or impossible to ascribe moral status in practice, and which generate interminable disagreements about the limits of moral concern. Consider problems involving *semantic indeterminacy*. Virtually every property that the individualist deems “status-conferring” is vague or ambiguous. Terms such as *sentience*, *consciousness*, *rationality*, or *agency*, have contested meanings not only within philosophy, but across psychology, neuroscience, and robotics (Gunkel 2014, 116). As David Gunkel puts it, these concepts are “undecided and considerably equivocal” (116). But without a clear understanding of what these concepts mean, it is hard to see what licenses the moral individualist to use them in ascribing moral status.

This criticism depends on two more general claims. First, that one cannot be justified in valuing *x*, without having a clear sense of what *x* is. And second, that moral individualism cannot be action-guiding unless one can determine which entities instantiate status-conferring properties. The first point has to do with indeterminacy involving a term’s intension, whereas the second has to do with its extension. While I think there is some plausibility to these objections, I worry that critics of moral individualism overplay their hand in taking them to be decisive. An individualist could concede that their theories are predicated on contested concepts, while denying that this poses unsurmountable problems. For example, Andreotta suggests that our intimate familiarity with first-personal conscious experiences licenses our use of the term “consciousness” in moral theorizing (Andreotta 2021, 25). We may not be able to define or explain the qualitative nature of phenomenal experience, but it is something that each of us is, presumably, well-acquainted with.

individualist should concede this. After all, their position—at least in many cases—is just that the absence of those features would give us good reason to *not* attribute moral consideration to SRs.

Semantic inferentialism provides a more promising (and to my knowledge, underexplored) way of developing this objection. Rather than frame the issue of semantic indeterminacy in terms of intension or extension, the critic might argue that the individualist relies on a cluster of concepts whose inferential relationships are poorly understood or obscure. Consider, for example, the relationship between *sentience* and having *interests*. Even if there were greater consensus about the definition or referent of these terms, one might still wonder about the entailment relations between them. Some philosophers contend that sentience is a necessary condition for having interests (DeGrazia 2015, 23). On this view, someone who claimed to be concerned about their plant's interests (e.g., in being watered) would be exhibiting a conceptual confusion. Other authors think that these notions can be held apart, such that one can intelligibly attribute interests to nonsentient entities or envision cases in which a sentient being did not have interests (Neely 2014, 98).[27]

In response, a moral individualist might accept that inferential connections between status-conferring concepts are messy, but insist that they can be given greater determinacy when contextualized. It may be pointless to ask *in general* whether having interests entails being sentient. What matters is that within a particular sociolinguistic practice, these relationships can be more clearly specified. In other words, it is possible to avoid the problem of inferential semantic indeterminacy by relativizing inferential relations to particular language games. Within a well-defined theoretical context, for example, one might be able to clearly specify the connections between concepts such as *sentience, interests,* or *harm*. But in doing so, one must concede that whatever normative conclusions follow from the arguments employing those

[27] Another example of the inferential indeterminacy would be the question of whether *consciousness* is conceptually separable from notions such as *intelligence* or *rationality*. Andreotta argues that these notions are independent of one another, such that it is possible to have an intelligent machine that is not phenomenally conscious (Andreotta 2021, 22-23). But one could envision someone who denied that this claim of independence.

concepts are only likely to appear compelling to participants in those practices. This response—as I shall discuss in greater detail below—is made available if one adopts a constrained version of individualism.

In addition to these conceptual issues, moral individualism also encounters a set of *epistemological problems*. Even if the individualist's central concepts could be clarified, it is not clear what evidence would warrant attributing these status-conferring properties to social robots. As David Gunkel puts it, "even if it were possible to define consciousness or come to some tentative agreement concerning its necessary and sufficient conditions, we still lack any credible and certain way to determine its actual presence in another. Because consciousness is a property attributed to 'other minds,' its presence or lack thereof requires access to something that is and remains fundamentally inaccessible" (Gunkel 2014, 117).[28]

While there may be borderline cases in which these epistemological problems arise, I doubt that, in general, individualists will find this argument convincing. As we have already seen, there are compelling argumentative strategies for attributing mentality to other entities on the basis of shared evolutionary history or biological similarities (Andreotta 2021).[29]

A more compelling objection, I think, is that moral individualism fails to provide a principled way of determining which intrinsic properties are morally relevant. Call this the *problem of relevance*. As Coeckelbergh puts it, since "[o]ur moral intuitions differ on what criteria are the relevant ones", questions about which ontological properties ground ascribing moral status suffer from an unavoidable indeterminacy (Coeckelbergh 2010, 212). Many people

[28] See also Coeckelbergh and Gunkel (2014, 718); Schwitzgebel and Garza (2015, 114).

[29] Moreover, there may also be independent reasons to be skeptical about the problem of other minds. For example, many philosophers regard Wittgenstein's so-called private language argument in *Philosophical Investigations* as administering the coup de grâce to this longstanding epistemological problem (Kenny 1973).

will readily concede that a dog's capacity to feel pain serves as a perfectly good reason not to kick them. But what about someone who does not share this intuition? One can easily imagine a person who is convinced that the concern for suffering is a distraction from what *really* matters morally—say, the possession of a soul. What is not clear is whether there is some way to properly adjudicate between competing first-order views of moral status, as moral individualism seems to require.[30]

One response, which I noted above, is to claim that any alternative to moral individualism will render ethics arbitrary (Schwitzgebel and Garza 2015). Unless there is some unique set of properties that grounds moral status universally, then our practices of showing moral concern will become subject to whim or fancy. But it is hard to see how *this* consideration bears on the problem of relevance. The spurner of suffering might be accused of many things, but arbitrariness does not seem to be one of them. Arbitrariness may, of course, arise in particular cases. For example, someone who decides, *without any justification*, to exclusively consider the interests of left-handed people would be acting arbitrarily. But for any property that the individualist deems morally arbitrary (e.g., left-handedness), it is always possible to imagine a background set of beliefs and desires against which that property *would* seem morally salient. It is, after all, the relative absence or presence of such beliefs and desires which explains why people disagree about whether possessing the soul matters morally.

The source of this problem, I think, is that moral individualism tends to run together two very different claims. The first is that moral agents ought to be consistent in their moral judgments and offer reasons for the properties they take to be morally relevant. If you believe

[30] This requirement follows from the fact that the moral individualist believes that the properties that ground moral status provide agent-neutral moral reasons.

that John's pain matters morally, then you ought to believe that Cindy's pain matters morally as well. The second idea is that some properties provide agent-neutral reasons whereas others do not. John's capacity to experience pain, for example, is supposed to provide such reasons, whereas his sinistrality is not. The problem of relevance is directed at this second assumption. It can be thought of as a general skepticism about the individualist's ability to deliver a convincing account of which properties do, and which do not, provide agent-neutral reasons.[31]

Some critics, whose views I shall discuss in the next section, take the problem of relevance to necessitate a radically different approach to moral status (Gunkel 2014; Coeckelbergh 2010; Coeckelbergh and Gunkel 2014). That is, they take the problem (in conjunction with the other issues discussed above) to undermine the very idea that *properties* can ground moral status. Ultimately, I shall I contend that these conclusions are too hasty. The problems discussed in this section are best addressed not through a total abandonment of moral individualism but by amending it. A *constrained individualism* holds that the question of which properties are morally relevant can be contextualized without rendering ethics "implausibly arbitrary". Adapting the central insight of metaethical constructivism, whether an agent regards a certain property as status conferring is a function of the practical identities they adopt. Put simply: relevance is a contextual notion, which depends on one's background beliefs and values constitutive of one's practical identity. This perspective preserves the individualist's idea that a moral patient's properties can factor into our reason-giving practices, while abandoning their claim that these reasons can be agent-neutral. This proposal will ultimately play an important role in my argument for a hybrid view of moral status, which I shall discuss in detail below.

[31] For Coeckelbergh, this skepticism targets the moral individualist's tendency to consider "entities at a distance" (2018, 148).

7.2.3 Moral Humanism and Social Robots

In this section, I consider a family of approaches to the moral status of social robots that explicitly sets itself in opposition to moral individualism. These views reject the idea that an entity's moral status depends on its properties. Rather, they share features which merit classifying them within the family of theories that I have been calling *moral humanism*.[32] First, these approaches are resolutely relational, taking seriously the idea that "an entity cannot be defined without refence to its relations" (Coeckelbergh 2014, 64). Second, they adopt a critical (in the Kantian sense) or transcendental perspective that prioritizes questions about the conditions of possibility of moral status ascription (Coeckelbergh and Gunkel 2014, 716). In doing so, these views are united in their rejection of the Cartesian notion that subjects and objects can be adequately conceived independently of one another. Humanists insist that subjects and objects (in this case moral agents and social robots) are mutually co-constituting, and they are eager to draw consequences from this view for moral status ascription.

Mark Coeckelbergh has developed the most detailed and compelling alternative to the individualist's "property-based" paradigm that dominates debates about the moral status of social robots (Coeckelbergh 2010; 2014; 2018). Like the moral humanists whom I discussed in Chapter Five, Coeckelbergh's *relational approach* embraces a Wittgensteinian view of language and culture. However, his work also draws from phenomenology, ecophilosophy, Marxism, and Deweyan pragmatism, thus broadening his focus beyond language and social practices.

[32] While "moral relationalism" would also be an apt label for this position, I refer to these positions as forms of *moral humanism* to emphasize their commonalities with the views outlined by Wittgensteinians such as Diamond and Crary. One advantage of this terminology is that it discourages thinking of this family of positions as advancing a competing *first-order* theory according to which moral status is simply *grounded* in relations—which could, in turn, provide agent neutral moral reasons. Rather, moral humanism is best thought of as engaged second-order questions about moral status ascription.

Like Diamond and other Wittgensteinian humanists, Coeckelbergh begins with the insight that entities cannot be adequately understood or defined apart from the social and natural relations in which they stand to other entities (Coeckelbergh 2014, 64).[33] This is not to substitute a relational ontology for a properties-based one—as this would just involve the substitution of one "dogmatic" approach for another (Coeckelbergh 2014, 65). Rather, Coeckelbergh contends that moral status is always ascribed within a complex socio-historical context, and that this context ought to be the subject of critical reflection. When it comes to the moral status of robots, adopting a relationalist approach would require one to know their "relations with other machines and with humans" and to understand how these entities are "naturally, materially, socially, and culturally embedded and constituted" (64). This requires taking seriously the thought that there is "no such thing as a robot-in-itself or thing-in-itself" (Coeckelbergh 2018, 150).

Coeckelbergh's relationalism departs from moral individualism in at least two respects. First, it begins with a thoroughgoing rejection of the Cartesian subject-object dichotomy. Rather than conceive of cognizing subjects and their objects as independent of one another, Coeckelbergh regards them "as mutually interdependent and mutually constituting" (2018, 149). Consequently, "[m]oral consideration is not only constructed in and by personal and social relations; our personal and social relations are at the same time constituted by other conditions and relations" (Coeckelbergh and Gunkel 2014, 724). On this view, moral status is not something that exists antecedently to and independently of social relations (or of any relations, for that matter). For Coeckelbergh, "moral standing is itself the outcome of the process of relation and interaction" (2018, 149). This is not to say that moral status ascription is simply a

[33] When it comes to social robots, he writes, "We need to contextualize moral standing, rather than study the entity as an atomistic curiosum in the anatomic theater of moral status science" (Coeckelbergh 2014, 64).

matter of fiat. Rather, we find ourselves "thrown" into a world of extant social relations, within which the limits of moral concern appear natural or fixed. As Coeckelbergh explains,

> when I, as a moral subject, "ascribe" moral status to an entity, I am not the first one to do so and the way I do it and the status I ascribe are probably already available in my society, my culture, and my language – more generally in what Wittgenstein would call my 'form of life'… Therefore, the question of moral standing is always connected to the question who is part of the moral community and what moral games are already played when and before I ask the question (149).

A relationalist approach seeks to inquire into these status-conferring relations by subjecting them to critical reflection.

Second, the fact that moral status ascription is always the outcome of context-dependent social relations suggests the need for a *critical* or *transcendental perspective*. Rather than inquire into the ontological grounds of moral status, a relational view interrogates the conditions of possibility for such ascription (Coeckelbergh 2014, 64). That is, it implies that "we need to reveal and criticize the social background of the question" (Coeckelbergh 2018 149). Consider, for example, how an entity's moral status is "partly constituted by the way we talk about it" (Coeckelbergh and Gunkel 2014, 724).[34] Echoing Cora Diamond, Coeckelbergh claims that our practices of *naming* have significant moral consequences, insofar as they function as a "way of demarcating the moral community" (725). For example, that people seldom consume their pets likely has a lot to do with how they relate to them. One way in which these relations are expressed is through "bestow[ing] singular proper names on an individual, and thereby individuated, animal" (725). As Coeckelbergh and Gunkel observe, "We call this specific dog 'Lassie' or that cat 'Mister Wiskers.' When an animal is named in this fashion, it often takes on

[34] In making this point, Coeckelbergh is not claiming that our relations to other entities are exhausted by our linguistic relations to them. As I mention below, he is also concerned with the moral implications of our affective, embodied, and other non-linguistic modes of relation to social robots.

face [*sic*] and is protected from abuse and killing. It becomes a 'pet', a 'family member', etc. rather than an 'animal'" (725).

From this perspective, moral individualism misses an important site for critical moral reflection and transformation. Naming practices are an important feature of our moral landscape that ought to be subject to moral reflection. But names are not properties. They are relational phenomena that cannot be adequately understood apart from the complex social interactions in which they are embedded.

As Coeckelbergh acknowledges, one potential drawback of a relational approach is its inability to provide clear, definite, positive practical recommendations (Coeckelbergh and Gunkel 2014, 728; Coeckelbergh 2018, 153). As he observes, "[T]his analysis of conditions of possibility for relations does not in itself advance a straightforward normative position" (728). Unlike the moral individualist, humanists like Coeckelbergh do not offer a set of necessary or sufficient conditions for attributing moral status. Their approach leaves things at a greater level of indeterminacy.

Nonetheless, one can glean some positive recommendations from Coeckelbergh's work. First, it motivates a strong form of anti-dogmatism about the moral status of social robots. As Coeckelbergh characterizes it, "a cautious, patient, and open attitude (and indeed character), then, can be said to constitute a meta-moral demand and a meta-virtue or moral-epistemic virtue" (Coeckelbergh 2018, 156). Given that moral status attributions are the outcome of relations and interactions that are themselves evolving, one ought to acknowledge that one's own attitudes and commitments concerning the limits of moral concern are likely to evolve as well.

A second positive feature of the relationalist turn is that we ought to take seriously the role of art in moral edification and reflection. Considering a number of performance pieces and installations querying the boundaries between humans and machines, Coeckelbergh writes:

> works of art such as these invite us to destabilize and critically question established meanings and borders, here to question the sharp border between machines and humans, or at least invite us to consider how in our imagination and feeling we already easily cross this border – whatever science or metaphysics may tell us (Coeckelbergh 2018, 155).

Finally, a relational approach requires that one take seriously "the phenomenology and experience of other entities such as robots" (2018, 149), by interrogating how these entities appear to us through our embodied, social relations with them (2018, 149; 2010, 214; 2014, 64). This requires contending with our emotional and affective responses to social robots, rather than viewing these phenomena as "appearances" or mistakes in need of correction.

Another advocate for what I am calling a humanist approach to the moral status of social robots is David Gunkel. In light of a set of perceived failures with moral individualism, Gunkel articulates a rival approach, which he calls "thinking otherwise" (Gunkel 2013). Drawing from the thought of Emmanuel Levinas, Gunkel proposes not so much an ethical theory (i.e., a normative or metaethical framework) as a "proto ethics" or "ethics of ethics" (Gunkel 2013, 127), which aims to "circumvent and deflect a lot of the difficulties that have traditionally tripped up moral thinking in general and efforts to address the moral status of the machine in particular" (126). I have already discussed some of these conceptual and epistemological worries in the previous section. Allow me to describe the central positive features of Gunkel's approach.

Like Diamond's humanism, Levinasian ethics inverts the individualist's starting point. Rather than begin by inquiring into an entity's status-conferring properties, Levinas maintains that "the ethical relationship, the exposure to the other, precedes the usual ontological decisions"

(Coeckelbergh and Gunkel 2014, 721). In other words, what matters morally "is not a set of predefined ontological properties", but rather "the intrusion of the face of the other that interrupts solitude, requires a response, and imposes a fundamental responsibility" (721).[35] This starting point represents at least two central departures from moral individualism.

First, whereas most moral individualists view the problem of other minds as an obstacle to be overcome, a Levinasian approach "affirms and acknowledges it as the basic condition of possibility for ethics" (Gunkel 2013, 126; Coeckelbergh and Gunkel 2014, 721). That is, those who accept the consciousness criterion as a necessary condition for moral status ascription owe an explanation of how we can ever be warranted in attributing consciousness to other entities, a path that continuously risks falling into skepticism given our lack of direct access to the mental states of others. By "thinking otherwise," Levinasians like Gunkel regard "the very condition of the ethical relationship" to be "an irreducible exposure to an other who always and already exceeds the boundaries of one's totalizing comprehension" (Gunkel 2013, 126).

A second feature of "thinking otherwise"—which is a corollary of rejecting the subject-object distinction—is that it dispenses with any firm distinction between agency and patiency. According to the common-sense philosophical picture that individualist's adopt, moral agents must deliberate about what they owe to moral patients. Not only are both sides of this relationship viewed as independent of one another, but it is assumed that they are constituted or "given" prior to ethical deliberation or action. The Levinasian picture jettisons this assumption. The self or the agent, as Gunkel puts it, "does not constitute some preexisting self-assured

[35] I take both Diamond and Levinasians like Gunkel to be making a similar transcendental argumentative move. Both are suggesting that, in order for moral individualism to get off the ground, some *X*, needs to be in place. For Diamond, that *X* is a broad, encultured linguistic competence which presupposes that certain concepts are already heavily normatively laden. For Levinas, that *X*, is a kind of phenomenological experience involving others.

condition that is situated before and as the cause of the subsequent relationship with an other… Rather, it becomes what it is as a byproduct of an uncontrolled and incomprehensible exposure to the face of the other that takes place prior to and in advance of any formulation of the self in terms of agency" (Gunkel 2013, 127). As one might expect, for Gunkel, these considerations do not suggest a straightforward answer to the question of whether robots can or should have rights. Rather, they represent a fundamental challenge to our default (i.e., individualist) understanding of ethics.[36]

7.2.4 Problems with Moral Humanism

These attempts to rethink the moral status of social robots beyond the individualist paradigm have encountered considerable resistance. The most common objection is that humanism amounts to an untenable form of relativism. That is, "taking the relational turn," or attempting to "think otherwise" renders impossible rational disagreement about moral status ascription, and leaves us without normative guidance. This objection is well expressed by Kestutis Mosakas, who writes:

> It seems that what the relational approach is fundamentally concerned with is our feelings and attitudes towards different entities, since that is what constitutes the basis of our relations; but without any central moral properties or guiding principles, it is difficult to see how this approach could genuinely help us in our moral decision-making without getting bogged down in a sea of relative judgements (Mosakas 2021, 434).

Similarly, Vincent Müller objects that "the core of the relational turn" is simply a "version of anything goes that dissolves the question of moral patiency to a random act of will" (Müller

[36] As Gunkel puts it, "The vindication of the rights of machines, therefore, is not simply a matter of extending moral consideration to one more historically excluded other, which would, in effect, leave the mechanism of moral philosophy in place, fully operational and unchallenged. Instead, the question concerning the 'rights of machines' makes a fundamental claim on ethics, requiring us to rethink the system of moral considerability all the way down" (Gunkel 2013, 130).

2021). The idea seems to be that abandoning the commitment to status-conferring properties entails the absurd conclusion that, "anything I happen to care about receives moral status" (Müller 2021). Thus, for humanists, no manner of ascribing moral status can be "better" or "worse" than any other.

A second, and closely related objection is that relational approaches fail to do justice to our moral intuitions, and therefore, fail to "track" morality (Mosakas 2021, 436). Consider, for example, the *Robinson Crusoe Problem*: if relationalism is true, then Robinson Crusoe (i.e., a person stranded on an island, standing in no social relations to others) would lack moral status, whereas Paro (i.e., a non-sentient baby seal-shaped robot used to treat dementia patients) would be morally considerable. For Mosakas, "that is a problem, because, in case of an ethical dilemma, it would seem that no one in their right mind should morally prioritize Paro over Crusoe" (435). Any theory that so flagrantly violates our moral intuitions ought to be rejected.[37]

How devastating are these objections to the relationalist's proposals? On my view, at best, the relativist challenge evinces the need for a more fully-elaborated account of how moral disagreements ought to be understood from a relational perspective. At worst, however, the claim that relationalism entails an objectional form of relativism rests on a false dichotomy and risks begging the question. To assume that, without status-conferring properties, one is left with an "anything goes" approach to moral status ascription overlooks the fact that there can be other

[37] I wonder about the coherence of this argument. First, it is unclear whether a Robinson Crusoe figure could even exist, given the simple fact that a human infant would not survive without the care of other humans. Perhaps the idea is that the figure stands in no *present* relations to others—such that anyone to whom they were once related has either permanently forgotten this fact or died (presumably, there would have to be no known records of Robinson Crusoe, since such knowledge would arguably generate some sort of social relation). Second, this argument presupposes a situation in which a moral agent must decide between furthering Crusoe's interests or Paro's interest. But this seems to entail that they would, by virtue of having to make this decision, enter into a social relation with Crusoe—even if only a very thin one. Thus, it seems implausible to say—at least by the relationalist's standards—that Crusoe would lack moral status.

sources of normative reasons, for example, and as we shall see, those stemming from a shared practical identity. It also assumes (implausibly, I think) that in order for a reason to be a moral reason, it must be one that *any* agent could be brought to accept. But this is, at best, contentious. Some relationships may serve as the basis for compelling—albeit limited—normative reasons that hold for others who fall outside of those relationships (Kittay 2005). But this does not mean that *any* set of relations will do so. What matters is that there are criteria that allow for reasoned disagreement about which relations serve as better or worse bases for our reasons.[38] Moreover, as we have seen, relationalists like Coeckelbergh do provide limited forms of normative guidance by recommending meta-normative values—such as open-mindedness—that ought to govern moral deliberation (Coeckelbergh 2018). These values can serve as a basis for rational critique of extant moral status ascriptions, and can go some way towards adjudicating between competing views about the limits of moral concern. This is obviously a less robust form of guidance than many individualists desire, but it begs all the important questions to assume (without argument) that some stronger form of normative guidance is necessarily required of a theory of moral status or that such guidance is even possible. Again, I believe that the charge of relativism is helpful *because* it challenges the relationalist to further articulate this aspect of their theory. But it risks circularity when it assumes that there is no avenue for a defensible relationalist position *unless there can be status-conferring properties*.

Setting aside the specter of relativism, I submit that there is a more serious problem. Relational approaches lack a *theory of error* which explains why properties-based approaches

[38] Plausibly, the fact that x and y are friends may serve as the basis for z's reasons for treating y a certain way. Whereas the fact that a and b are standing six feet apart may not generate such reasons for c. The difference is explained by the shared understanding and importance of friendship.

seem to capture so many people's moral intuitions.[39] Undeniably, we often *do* appeal to an entity's properties when justifying our treatment of it. There is something plausible about the idea that our moral concern for others is, in part, grounded in their properties or attributes. *Because she can feel pain* is, on the face of it, a reasonable response to the question of why one ought not to pull the cat's tail. Although proponents of the relational turn raise compelling conceptual, epistemic, and practical problems for a *general theory* of moral status limited exclusively to intrinsic properties, they fail to explain moral individualism's intuitive appeal.

Therefore, either the humanist needs to explain these intuitions away, or they must find a way of accommodating them. The proposal developed in the remainder of this chapter opts for the latter option. Humanist approaches go too far in completely eschewing status-conferring properties from moral practice; and, consequently they fail to appreciate the possibility of an attenuated and more narrowly circumscribed version of the properties view. The *constrained individualism* which I introduced above, affords properties a role within reason-giving, while denying that they can be understood independently of their relations to our background beliefs and values. Put another way, a *constrained humanism* accepts that properties can be morally relevant, but insists that relevance is a relative notion. This relativity, I want to suggest, can be understood as relative to a concept introduced in the previous chapter—the notion of a practical identity.

[39] That is to say, given that we at least *appear* to be justified in appealing to an entity's non-moral properties when justifying our treatment of it, relational accounts (and moral humanism, more generally) owe an explanation, not only of *why* these justifications are mistaken, but of *how* such a justificatory error became so ubiquitous.

7.3 Practical Identity Proliferation: A Hybrid Account of Moral Status

One upshot of the discussion so far is that there is no consensus about how to best inquire into moral status of social robots. Moral individualism captures the widely held intuition that properties such as sentience, intelligence, or sapience, are morally relevant; and it translates this intuition into a simple, compelling argument that provides normative guidance. But it fails to consider the social and historical contexts in which moral status is ascribed and suffers from conceptual and epistemological problems. Moral humanism seems better positioned to make up for these shortcomings through its emphasis on context and the complexity of our conceptual practices; but its limited normative guidance and failure to afford place to status-conferring properties within our reason-giving practices constitute serious drawbacks.

In the remainder of this chapter, I shall advance and defend a hybrid approach to moral status ascription incorporating elements of both individualism and humanism. This position includes two components. On the one hand, it involves a *first-order* view, which claims that we should adopt a pluralistic, fallibilistic, open-ended stance towards the properties and relations that ground moral status. On the other hand, it involves a *second-order* view according to which we ought to adopt *both* individualistic and humanistic deliberative strategies when ascribing moral status. But given the apparent incompatibility of individualism and humanism this might seem untenable. What is needed, therefore, is (1) an argument that can demonstrate the continued importance of both sets of strategies, and (2) a blueprint outlining how they can operate together in practice. The pragmatic genealogy developed in the previous chapter delivers on both scores.

Pragmatic genealogies explain the existence and shape of our conceptual practices in terms of their ability to satisfy our needs or interests. As I argued in Chapter Four, this explanatory structure permits a kind of normative purchase. A genealogy is vindicatory when it demonstrates (in some non-trivial way) that a target practice responds to problems that we do as matter of fact continue to face, or when that response remains stable upon reflection (e.g., in light of its compatibility with other intrinsic values or concerns that we hold). Alternatively, a genealogy may undermine or subvert our commitment to a target practice by revealing it to be responsive to interests that we no longer should care to endorse, or as embodying patterns of response that we take to be defective in some way, perhaps in light of available alternatives, or because they cannot remain stable under reflection.

Whatever its normative ambitions, however, any pragmatic genealogy will be constrained by the *agent-relativity* of the needs it ascribes. As Queloz puts it:

> The idea of a need is correlative with the idea of a serious harm that one will incur if the need is not satisfied, and that idea of a serious harm is in turn correlative with culturally conditioned conceptions of human life and flourishing (Queloz 2021, 237).

That is, insofar as it makes explanatory use of generic needs, interests, or purposes, a pragmatic genealogy will always rely—either tacitly or overtly—on some conception of an agent for whom those needs, purposes, or interests can be said to obtain (Queloz 2021, 238). Thus, the genealogist's explanatory and normative conclusions will always be conditional. In most cases, the relativity of needs to conceptions of agency will be uncontroversial. Nobody would deny that our need to drink water is relative to our biological makeup. Arguably, however, this constraint will be most strongly felt when one looks to a pragmatic genealogy's secondary elaboration. The more historically localized that practices under consideration become, the more likely one is to encounter competing conceptions of what an agent is. Consider, for instance, the relatively uncontroversial claim that—at some very high level of abstraction—our *generic* need for gathering and sharing information is predicated on the fact that we are epistemically limited beings. But as one considers more de-idealized or localized practices, there is room for disagreement about what counts as a need *qua* gathering and sharing information. For example, in a technologically advanced society such as ours, does literacy count as a human need? What about access to the internet? It is hard to see how someone could provide convincing answers to these questions without falling back on some culturally conditioned—and likely contested—conception of what an agent is. In part, this variability is due to the emergence and increased complexity of cultural affordances, but it is also because of diverging conceptions of "human life and flourishing."

Queloz's discussion of Bernard Williams's treatment of this problem is illuminating and is worth examining in detail because it can serve as a model for my own attempt to articulate a conception of agency that can throw light on the first and second-order questions about moral status ascription under consideration.

In a late essay entitled, "From Freedom to Liberty: The Construction of a Political Value," Williams offers an account of the value of liberty and its central role within political liberalism (Williams 2005). This "vindicatory genealogy of liberty" (Queloz 2021, 239) follows the contours of the two-stage explanatory model that I discussed in Chapter Four. In what can be considered a state-of-nature model, Williams contends that—in some generic, abstract sense—any society will have a need for "primitive freedom" consisting of the "simple idea of being unobstructed in doing what you want by some form of humanly imposed coercion" (Williams 2005, 79). As Queloz notes, this basic conception of freedom is connected to *any* notion of agency, in the sense that it is difficult to imagine an *agent* who remained entirely indifferent to violations of their primitive freedom.

For Williams, *primitive freedom* is a pre-political notion whereas *liberty*—the sense of freedom that interests Williams—is an inherently political concept that is tethered to historically specific practices and institutions. For this reason, one cannot appeal to the former notion—i.e., the thin "primitive" concept of freedom—in order to justify preferring a liberal notion of liberty over any other configuration of political values and institutions. The need for primitive freedom is one that could be equally well satisfied by a variety of forms of social organization. And this points to another problem. The more local, liberal conception of liberty is tied to its own historically contingent conception of an agent, thereby threatening a kind of circularity. As Queloz observes, "to justify the liberal order in terms of a coeval conception of the agent that only liberals accept is mere self-congratulation" (Queloz 2021, 239). In other words, it will not do to justify political liberty by declaring it a human need, because the very claim that it *is* a need borrows its plausibility from a particular conception of agency which "fits the liberal order because it emerged alongside it" (239). What is needed is a way for defenders of liberalism to achieve "reasonable confidence" in the value of political liberty which is not presupposed by liberalism itself. As Queloz explains, this "requires achieving a vindicatory reflective understanding of liberty as a value: an understanding, among other things, of why we have it, what needs it answers to, and whether it is right for us given our circumstances" (239). Moreover,

> It must include a reflective understanding of the basic concerns to which a more generic notion of freedom answers, and of why the socio-historical elaboration of the notion of freedom we happen to have is adequate to our socio-historical elaboration of those basic concerns (239).

On Queloz's reading, this is precisely what Williams aims to establish through de-idealizing the generic idea of primitive freedom. The goal is to "construct" a political conception of liberty that is responsive to more localized practical demands (Williams 2005, 84). Very roughly, the story goes something like this: primitive freedom engenders inevitable interpersonal conflicts that require some public method of resolution. As individuals exercise their primitive freedom, they are bound to encounter situations in which their own freedom-demands run into conflict with the freedom-demands of others. This, Williams thinks, will require some way of distinguishing between legitimate and illegitimate uses of public authority for resolving such conflicts. Hence the need for "legitimation accounts" that enable people to accept public forms of adjudication, and that remain endorsable upon reflection.[40] But what might such legitimation stories look like? That depends. For Williams, over the past few centuries the criteria governing the acceptability of legitimation stories have undergone significant transformations. While appeals to the divine right of kings may have satisfied our political forebearers, such models of political legitimation are largely ineffective today (Williams 2005, 95). In part, this is because modernity has brought with it a high degree of "historical self-consciousness" which has undone many of the traditional forms of political legitimation (Queloz 241). As Queloz explains,

> Under conditions of modernity, truthful inquiry and self-consciousness have eroded many of the myths, narratives, and Whiggish histories that formed the stuff of past legitimation stories, leaving us with less material for our legitimation stories; and once these sources of legitimation have fallen away, there is a stronger presumption in favour of citizens' freedom to do what they decidedly want (240-1).

It is in light of this relatively recent set of sociohistorical circumstances that *political liberty* as a distinctive value begins to serve an important purpose. Several centuries ago, perhaps, the need to have *some* kind of legitimation story could have been satisfied by a set of concepts and ideals (e.g., the divine right of kings) that are no longer functional today. As Queloz emphasizes, it is the fact that *we* are left with "less material for our legitimation stories" that generates a kind of political *need* to which the value of *political liberty* serves as a response.

[40] See also Williams (2005, Chapter One).

Williams's discussion indicates why a conception of liberty is reasonable for certain kinds of agents—namely those who must navigate a set of historically local political problems within liberal democracies. Analogously, I want to suggest that a hybrid approach to moral status ascription is reasonable *for certain kinds of agents*, namely, those who encounter practical problems arising within societies such as ours.

7.3.1 Practical Identity Proliferation and the Need for a Hybrid Approach to SR Rights

The expression *practical identity proliferation* denotes a set of social conditions under which individuals find themselves adopting and navigating an ever-growing swath of practical identities. Throughout our lives, we members of contemporary liberal democratic societies have the potential to be so many kinds of people, to identify with so many groups, to occupy so many social roles. The diversity of practical identities that present themselves as what William James called "live options," seems far greater today than it has at any previous point in history.

This poses challenges insofar as our identities matter to us. They shape our sense of self-worth and provide a source of our values and motivations. As Korsgaard puts it, "Our conceptions of our practical identity govern our choice of actions, for to value yourself in a certain role or under a certain description is at the same time to find it worthwhile to do certain acts for the sake of certain ends, and impossible, even unthinkable, to do others" (20). This idea is, however, not limited to metaethical constructivism, but has been developed in considerable detail within other areas of philosophy—notably feminist philosophy (Lindemann 2019, chapter 4) and pragmatism (Rorty 1989). It is finds empirical support from social identity theory and self-categorization theory (Jenkins 2014).

Without claiming to offer an exhaustive theory of practical identity proliferation (PIP), I want to suggest that we can separate analytically three of its core features. First, PIP involves an increase in the *variety* of practical identities available within a social context. This might include identities afforded by membership in social groups, political associations, professional affiliations, religious identities, identities involving one's gender or sexual orientation, among others. Second, PIP typically involves a high degree of *self-awareness* accompanying identities. That is, as people are confronted with a greater number of options, and as they attempt to juggle the demands of those identities they have adopted, they are more likely to reflect upon the meaning and implication of those identities. A third core feature is that identity proliferation is accompanied by an increased *contestability* of practical identities. The process through which new social identities emerge involves disagreement and collective deliberation.

Whatever its historical origins, the phenomenon of practical identity proliferation is difficult to deny. [41] We have become the kinds of agents who must navigate the complexities and

[41] Although my argument only requires that practical identity proliferation is, in fact, a pervasive feature of the contemporary world, I also believe the stronger claim that the extent to which this is the case is far greater than it has been at previous points in human history. To motivate this stronger hypothesis, allow me briefly consider some hypotheses about the factors driving this phenomenon. These proposals are not offered as monocausal historical explanations of practical identity proliferation; rather, the idea is that, when taken together, they lend support to the claim that it is has increasingly become widespread and they help bring its contours into sharper focus.

Consider first, how *secularization* has contributed to practical identity proliferation. As people in western societies gradually abandoned a shared Christian worldview, they needed to find additional sources of meaning and projects of self-realization and moral understanding. It seems reasonable to suppose that this general process of secularization has led to a diversification of social identification.

Another force that has arguable played a role in practical identity proliferation is *bureaucratization*. The emergence of new public social roles afforded opportunities for previously unimaginable forms of social identity. Similarly, and more recently, changes in the trajectory of an average career indicate that, over time, people are taking on increasingly more roles in their professional lives.

Civil rights movements of the 20th century represent a twofold explanatory factor. On the one hand, involvement in social movements is a significant source of meaning and fulfilment for many people. When someone participates in the fight for a political cause they often take on a practical identity that they shared and negotiate with others. On the other hand, the loosening of social stigmas and prejudices brought about by civil rights movements has arguably been accompanied by a proliferation of identities associated with sexual orientation and gender expression.

Finally, *digital technologies* such as the internet and social media have given rise to new subcultures that simply would not have been possible in the past. In some cases, the shared use of a technology itself has been a site dynamics of a world of proliferating social roles. With this idea in mind, I can now present an argument for a hybrid approach to moral status whose structure resembles Williams's vindication of political liberty. But whereas Williams elucidates the point of a particular conception of political freedom, I aim to elucidate the point of (and therefore, to legitimate) a set of deliberative strategies for attributing moral status.

My argument aims to vindicate a hybrid view of both second-order and first-order questions considered above. Allow me to unpack the latter claim before presenting a similar argument for first-order questions. Put schematically, the argument goes like this:

> *Vindication of second-order hybrid approach*: as agents who must navigate a social world characterized by practical identity proliferation, we encounter a set of distinct practical problems. Moral individualism and moral humanism describe strategies for resolving these problems, neither of which is independently sufficient for doing so. Therefore, we are justified in accepting both sets of deliberative strategies.

This argument builds on two results of the previous chapter: first, that moral status ascription is tethered to the practical identities one adopts; and second, that moral individualism and humanism track diverging deliberative strategies for managing practical identity conflicts. The question that I am considering is this: given a world of proliferating practical identities, which of these two deliberative strategies ought we accept? The answer I want to defend is that we require both. When it comes to navigating changes and conflicts amongst our proliferating practical identities, sometimes individualist strategies outperform humanist strategies, in other cases the opposite is true. Allow me to explain.

By appealing to the importance of status-conferring properties, moral individualism facilitates rational disagreement about the limits of moral concern. Crucially, it even aims to do

of formation for practical identities (e.g., the gamer). The development and integration of social robots into our lives is only likely to accelerate and intensify this process of identity proliferation.

so when the parties to such disagreement have radically different practical identities. That is, the appeal to status-conferring properties is meant to generate reasons for action while abstracting away from the practical identities in question. By distilling questions of the basis of moral status down to a simplified—and supposedly agent-neutral—set of properties, the moral individualist aims to build a bridge between interlocutors whose practical identities would otherwise leave them with conflicting (and even irreconcilable) normative orientations.

In this respect, moral individualism exemplifies what C. Thi Nuygen calls a *value capture* strategy (Nguyen 2021, 422-3). Value capture occurs when:

(1) Our values concerning *x* are complex and difficult to express.
(2) We find ourselves in a social or institutional context which simplifies (often in quantified form) those values, and presents them back to us.
(3) Those simplified values take over our motivations and deliberation.

So often, in taking up a practical identity, one finds oneself confronted with a welter of ideals and values that are not readily articulable. To occupy a social role is to adopt attitudes and modes of responsiveness that may seem natural "from the inside," even when the motivations for those attitudes or responses are difficult to explain or justify. In particular, it can often be challenging to specify why one's practical identity calls one to direct moral attention and concern towards certain moral patients rather than others. For example, someone who thinks of themselves as an environmentalist may find certain courses of action appealing or even required of them even without being able to articulate precisely why this is the case. They may, for instance, feel a sense of direct obligation to elements of the natural world, say, without being able to precisely articulate why. Perhaps they might appeal to their love of nature, the importance of conservation, or some vague desire to promote biodiversity—perhaps they may simply appeal to the fact that they care about the environment. The individualist's deliberative strategy is to effectively cut through this morass of values and sentiments, thereby delivering a kind of *value clarity* about the scope of one's obligations and the reasons underlying them.[42]

When it comes to moral status attribution, this process of value simplification is useful for agents who experience a world of proliferating practical identities. Often, we encounter situations in which we want individuals to align (at least in part) in their moral attitudes without expecting them to adopt the same practical identities. One way to do this is by abstracting from the variety of practical identities involved in a social practice by making salient some simplified—and hopefully shared—values that can guide decision-making. In this respect, moral individualism represents a means of solving moral coordination problems that we tend to find in public institutions. That is, many public institutions require an alignment of attitudes (especially concerning other participants in those institutions) without requiring a "leveling off" of participants' background beliefs and desires.

For example, consider how successful moral individualism has been within research ethics contexts. The recommendation that sentience serves as a morally relevant property is a drastic simplification of the values that a researcher might bring to bear on the question of how she ought to treat her research subjects. But the point of the strategy—and the reason it is so effective—is that research contexts are precisely those kinds of cases in which participants in a shared practice must converge on certain attitudes and conduct while maintaining a wide range of background practical identities. Focusing on the simplified value of sentience enables policies

[42] One way in which moral individualists have offered a value capture strategy to environmentalists is through various accounts of *biocentrism*, which maintain that all living things have intrinsic value. This provides a way of simplifying the complex values and motivations underlying an environmentalist worldview in order to facilitate deliberation and the public justification of their projects.

that are, to some extent, generalizable, and therefore, acceptable to a wide range of agents whose fundamental normative outlooks may differ radically.

Moral individualism is also an effective deliberative strategy within legal contexts. Consider how individualism could make possible the project of justifying legal rights to robots. Granting robots legal rights requires a broad consensus about how they ought to be treated with respect to the law. Crucially, this consensus must be secured even amongst those with wildly diverging practical identities. Appealing to status-conferring properties serves as a kind of "common moral currency" in these contexts.[43] In doing so, moral individualism captures our values, then feeds them back to us in simplified form in order to facilitate participation within public institutions.

Nonetheless the moral individualist's value-capture strategy is of scant use in other deliberative settings. Individualism captures and simplifies our values in the service of social cohesion. But so often what is needed are tools that can advance deliberation while preserving, or even increasing our sensitivity to the complexity of our values and the indeterminacy inherent to any morally problematic situation. Moral humanism offers such a normative orientation—what one might call a *complexity preserving* form of deliberation. Hence, its focus on questions such as: what are the historical and social conditions under which moral status ascriptions take place? How do our linguistic practices, such as naming, shape the boundaries of our moral communities? How do other entities appear to us in the course of our moral experience? These are not necessarily the kinds of questions that one asks when one is trying to secure moral agreement with others. Rather, they are the kinds of questions someone tends to ask themselves

[43] This is evident in the debate concerning the European Parliament's proposal to recognize "artificial personhood" as a legal category.

when deciding which sort of person they want to be, or that social groups tend to consider when confronted with the task of articulating their self-image. The point of these deliberative activities is not to simplify one's values to facilitate reasoned disagreement amongst those with diverse interests, but to exercise one's imagination in order to gain a better sense of what one takes to be possible and important.

It is true that engaging in these sorts of deliberative processes is unlikely to yield *arguments* that can resolve disputes about moral status ascription, or to provide a set of agent-neutral *reasons* for circumscribing the limits of moral concern in a particular way. But to find fault with the moral humanist for failing to produce such arguments or reasons is to miss their point. Moral humanism is not directed at those dimensions of moral practices in which we are after simplified value clarity; rather, it is directed at those (perhaps rare) exercises of imagination in which we are ultimately led to rethink and transform who we are, and thus, to rethink and transform our intuitions about the scope of moral concern. Moral humanism is, therefore, oriented to the possibility of radical moral transformation.

I am claiming that both sets of deliberative strategies are useful within a world of practical identity proliferation. On the one hand, I have argued that moral individualism is best suited for public contexts in which we need to simplify our values to solve coordination problems. However, it is important to recognize individualism's inherently conservative nature. It extracts morally relevant properties by abstracting from the practical identities within which those properties are initially understood as salient. But it cannot lead one to accept a given property as morally relevant when one does not do so antecedently. On the other hand, moral humanism, has the potential to expand or transform our values.[44] It is only through engaging the imagination that we are likely to regard as morally relevant features that were once seen as unimportant. Nonetheless, one should not expect moral humanism to be of much use for securing rational agreement. But this is just to say that the two deliberative strategies answer to very different practical problems that agents like us are likely to face.

So far, I have been arguing for a *second-order hybridism* that embraces both moral individualism and humanism. On my view, however, a similar set of considerations motivates a hybrid stance towards first-order questions about which properties should be considered morally relevant.

> *First-order hybrid approach*: When attributing moral status, we ought to take a fallibilistic, open-ended attitude towards those properties we currently regard as morally relevant.

I claim that the reasonableness of this position follows from two plausible assumptions, both of which have been expounded in this chapter. First, that we are the kind of agents who must navigate the challenges imposed by practical identity proliferation. And second, that a property's moral relevance is a function of the practical identities one adopts. Taken together, these propositions imply that the emergence of new practical identities or the transformation of extant identities can change which properties one takes to be morally relevant. If this is the case, then given that one's own practical identities are liable to change over time, one ought to accept that one may ultimately revise one's attitudes towards the properties one currently takes to be status-

[44] As Coeckelbergh and Gunkel put it, ethical approaches that "deploy and endorse a properties approach to moral status ascription will fail to achieve any real moral progress" (721). Whether or not they are correct, of course, depends on the conception of progress they have in mind. I have argued in the previous chapter that humanists and individualists are tracking different conceptions of moral progress.

conferring. The proper meta-normative stance to take *viz* moral status ascription is, therefore, one of open-mindedness and fallibility.[45]

Of course, this hybrid view require that, at least to some extent, moral individualism and humanism can be reconciled. As I have argued throughout this chapter, my view is that this reconciliation is possible so long as one is willing to accept constraints on both positions. Crucially, it requires adopting a form of constrained individualism, which abandons the idea that there can be status-conferring properties that swing free of *any* practical identity. This preserves the individualist's insight that status-conferring properties play a legitimate role within our moral reason-giving practices. But the proposal contextualizes this insight. An entity's properties can be morally relevant, but they are always relevant for certain kinds of agents. This amendment circumvents the problem of relevance by eschewing the implausible claim that status-conferring properties wear their relevance on their sleeve. This, of course, is a substantial revision to moral individualism, and in order for a hybrid approach to get off the ground it needs to be matched with an equally substantial concession from moral humanism, namely, that status-conferring properties do show up "from the inside" so to speak, as morally relevant. That is, the humanist needs to acknowledge a place for a properties-based view within our moral lives.

It is important not to lose sight of the conditional nature of this argument. In particular, I have not claimed that *everyone* has a good reason to go in for a hybrid account of moral status. A form of life in which there was relative homogeneity amongst practical identities might not need an hybrid approach. Without a plurality of background values, it may be unnecessary to employ

[45] Consider, for example, how attitudes towards the importance of a soul have gradually changed over the past century or so. In part, this is because people are increasingly turning away from religion for moral guidance; but it is also due to changing self-conceptions of what religious identities demand. My claim is that these latter changes are responsible for the waning relevance of "possessing a soul" for moral status attribution.

a value capture strategy (especially when designing one's public institutions). Similarly, a group of people who simply did not care about transforming their values might find that they have no use for the deliberative strategies marked by moral humanism. Rather, my argument depends on the claim that *we* are the kinds of agents who require both sets of strategies. Given the fact that we are agents whose practical identities are proliferating, we have good reason to adopt a pluralistic stance towards first-order questions about which properties are status-conferring as well as to recognize the importance of both value-capture and complexity-preserving deliberative methods. This, as I have claimed, requires making substantive revisions to both frameworks. In particular, *pace* standard forms of moral individualism it requires relativizing the relevance of status-conferring properties to certain practical identities. *Pace* humanism, we ought to accept the idea that certain properties are morally relevant—but in a way that is much more constrained than is typically thought.

7.3.2 Applying the Hybrid Approach

Having outlined the basic idea behind a hybrid approach to moral status ascription, I can now explain how it applies to debates about robot rights.

Generally speaking, one of the most obvious implications of the view is that it advises against speaking about the moral status of social robots *in general*. Thus, it calls into question standard versions of moral individualism that attempt to specify a set of necessary conditions that a social robot would need to meet in order to have moral status *no matter the context.* On a hybrid approach, these kinds of questions always require careful contextualization. To better understand the implications of my view, therefore, one would need to identify the practical contexts in which moral individualism and moral humanism's deliberative strategies are called for. I have already offered a general characterization of how this might look. But allow me to spell this out in terms of the case of social robots.

First, one feature of the constrained form of individualism I recommend is its *domain specificity*. Rather than ask which properties ground moral status in general, one should begin by looking to the contexts where the individualist's value-simplification strategy would be most appropriate. On my view, the most salient contexts are those in which the following two conditions hold:

(i) Moral agents with potentially radically different practical identities will need to interact with social robots.

(ii) These situations require agents to coordinate their attitudes concerning the kinds of treatment of robots deemed acceptable and unacceptable.

Intuitively, one could envision these conditions being met as advanced social robots are eventually employed in domains such as healthcare, education, industry, entertainment, and the military—just to name a few examples. These are precisely the kinds of domains in which it is necessary to coordinate the behavior of agents whose background values beliefs and values often exhibit significant variability.

In these cases, a constrained form of individualism might begin by identifying the stakeholders involved in the situation and considering the kinds of practical identities likely to be involved. One strategy for solving the coordination problem just mentioned would be to identify properties that are likely to be seen as morally salient from the perspective of the relevant stakeholders and to make moral status ascription on the basis of those.

For example, consider a not-so-distant scenario in which social robots are integrated into healthcare systems, performing patient care in hospitals and homes. With advances in natural language processing, it is not difficult to imagine a sophisticated robot who monitored and responded to patients' health needs, interacted with them socially and in ways that demonstrated autonomy and even emotional intelligence. In scenarios such as this, one could also envision patients and their families coming to form emotional bonds with these healthcare robots—even coming to display paradigmatically *moral* attitudes towards these entities such as gratitude and trust. If, in these kinds of cases, participants in the healthcare system began raising questions about how they and others ought to treat these social robots, then a hybrid approach to moral status would recommend relying on a constrained individualist strategy as a guide to policy making. This would involve considering various candidate properties as morally relevant *within the context*, and proposing policy measures concerning the treatment of SRs on the basis of those properties. For example, capacities for autonomy, emotional and psychological complexity might be considered status-conferring in these contexts.

One way in which this constrained form of individualism (which constitutes one dimension of a hybrid approach to moral status) differs substantially from its unconstrained alternative, is that it does not begin by assuming, at the outset, which properties are status-conferring. Instead, my view begins by looking to contexts in which people antecedently face moral coordination problems and then attempts to build consensus from there. Another distinctive feature of this strategy is that it encourages us to regard the properties we *do* currently take to warrant moral consideration as open-ended and revisable.

Whereas constrained individualism will be most applicable to contexts satisfying the two conditions mentioned above, moral humanism is most applicable to what can be understood as contexts of moral self-transformation and social criticism. Consider for example, how the moral humanist's deliberative strategies might be employed to critically examine our evolving relationships to social robots. As we have seen, these strategies include phenomenological reflection (i.e., attention to the qualitative nature of our experience interacting with social robots), cultural critique, and transcendental reflection (e.g., reflection how social institutions and language make possible certain modes of interaction with social robots).

As social robots increasingly come to inhabit our shared social world, a constrained version of moral humanism will be especially useful for facilitating reflection on our shared practical identities with those machines. For the most part, the practical identities we currently adopt—be it our familial roles, professions, memberships in various organizations, religious affiliations, and so on—are ones we share with other humans. But as social robots develop complex capacities that enable them to *participate* in social practices, it is easy to imagine near-future cases in which we would begin to ask whether *they* could be said to occupy these shared identities as well. The healthcare robots just mentioned present a case in point. Their integration into medical practices would almost certainly raise questions about what it means to be a healthcare provider. Might practitioners someday regard the intelligent machines with whom they increasingly interact and cooperate with as "fellow surgeons"? On my view, these questions are not ones that are best answered by attempting to *simplify* our values in the service of cooperation with others. Rather, these kinds of question require critical and imaginative reflection—that is to say, they demand the sort of complexity-preserving deliberative strategies found in moral humanism.

On the constrained version of moral humanism that I endorse as part of my hybrid view, an important resource for critical reflection on our possible *shared practical identities with social robots* would be the production and enjoyment of art, literature, and film. These mediums challenge us to rethink our existing practical identities but also the possibilities of future identities. In doing so, they can help us reassess the meaning and relevance of the properties and relations we *do* currently see as salient and important. For example, films like *Her*, or *Ex Machina* challenge us to rethink the meaning of notions such as *intelligence*, *friendship*, *suffering*, *agency*, and *trust* through their depictions of human-machine interaction. Art and literature can extend the use of our concepts to new situations, thereby affecting a kind of moral reorientation. This is why constrained humanism is well-suited to these contexts of moral self-transformation (and social criticism). Crucially, on the hybrid approach that I recommend, these projects of moral self-transformation and social criticism encouraged by moral humanism may ultimately lead to revisions in the individualist-oriented policy-making within our shared institutions.

BIBLIOGRAPHY

Aikin, Scott, and Michael Hodges. 2018. Expressivism, Moral Judgment, and Disagreement: A Jamesian Program. *The Journal of Speculative Philosophy* 32 (4): 628-56.

Aikin Scott, and Robert Talisse. 2011. Three Challenges to Jamesian Ethics. *William James Studies* 6: 3-9.

Allen, Barry. 2003. Another New Nietzsche. *History and Theory* 42 (3): 363-77.

Altham, J. E. J. 1995. Reflection and Confidence. In *Essays on the Ethical Philosophy of Bernard Williams*, ed. J.E.J Altham and Harrison Ross. Cambridge: Cambridge University Press.

Anderson, Elizabeth. 2004. Animal Rights and the Values of Nonhuman Life. In *Animal Rights: Current Debates and New Directions*, ed. Cass Sunstein and Martha Nussbaum. Oxford: Oxford University Press.

———. 2021. “How to be a Pragmatist.” In *The Routledge Handbook of Practical Reason*, ed. Ruth Chang and Kurt Sylvain. New York: Routledge.

Andreotta, Adam J. 2021. The hard problem of AI rights. *AI & Society* 36 (1): 19–32.

Appiah, Anthony Kwame.2006. *Cosmopolitanism: Ethics in a World of Strangers*. New York: W. W. Norton & Company.

Aristotle. 2016. *Metaphysics*. Trans. C. D. C. Reeve. Indianapolis: Hackett Publishing Company

Ayer, A. J. 1952. *Language, Truth and Logic.* New York: Dover.

Bacon, Michael. 2012. *Pragmatism: An Introduction*. Cambridge: Polity.

Bentham, J. 1789/2000. *An Introduction to the Principles of Morals and Legislation.* Kitchener: Batoche Books

Bernstein, Mark H. 1998. *On Moral Considerability*. Oxford: Oxford University Press.

Blackburn, Simon, and Bernard Williams. 1986. Making Ends Meet. In *Philosophical Books* 27 (4): 193-203.

Blackburn, Simon. 1984. *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Clarendon Press.

———. 1998a. *Ruling Passions: A Theory of Practical Reasoning*. New York: Oxford University Press.

———.1998b. Wittgenstein, Wright, Rorty and Minimalism. *Mind* 107 (425): 157–81.

———. 2009. The Landscapes of Pragmatism. *Teorema* 28 (3): 31–48.

———. 2013. Pragmatism: All or Some? In *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press.

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bostrom, Nick., and Eliezer Yudkowsky. 2014. The Ethics of Artificial Intelligence. *The Cambridge Handbook of Artificial Intelligence*, ed. Keith Frankish and William M. Ramsey. Cambridge: Cambridge University Press.

Brandom, Robert. 1994. *Making It Explicit*. Cambridge: Harvard University Press.

———. 2000a. *Articulating Reasons*. Cambridge: Harvard University Press.

———. 2000b. Vocabularies of Pragmatism: Synthesizing Naturalism and Historicism. In *Rorty and His Critics*, ed. Robert Brandom. Oxford and Cambridge: Blackwell.

———. 2009. *Reason in Philosophy: Animating Ideas*. Cambridge: Harvard University Press

———. 2013a. An Arc of Thought: From Rorty's Eliminative Materialism to His Pragmatism. In *Richard Rorty: From Pragmatist Philosophy to Cultural Politics*, ed. Alexander Groschner, Colin Koopman, and Mike Sandbothe. New York: Bloomsbury.

———. 2013b. Global Anti-Representationalism? In *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press.

Brey, Philip. 2008. Do We Have Moral Duties Towards Information Objects? *Ethics and Information Technology* 10: 109–14.

Brooks, Rodney. 2000. Will robots rise up and demand their rights? *Time*, June 19. https://content.time.com/time/subscriber/article/0,33009,997274,00.html

Bryson, Joanna J. 2009. Robots Should be Slaves. In *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, ed. Yorick Wilks. Philadelphia: John Benjamins Publishing Company.

Buchanan, Allen. 2009. Moral Status and Human Enhancement. *Philosophy & Public Affairs* 37 (4): 346-81.

Buchanan, Allen, and Russell Powell. 2016. Toward a Naturalistic Theory of Moral Progress. *Ethics* 126: 983–1014.

Cantrell, Michael A. 2013. William James's Transcendental Theological Voluntarism: A Reading of "The Moral Philosopher and The Moral Life". *William James Studies* 10: 1–10.

Capps, John. 2018. From Global Expressivism to Global Pragmatism. *Metaphilosophy* 49 (1-2): 71-89.

Cappuccio, Massimiliano L., Anco Peeters, and William McDonald. 2019. Sympathy for Dolores: Moral Consideration for Robots Based on Virtue and Recognition. *Philosophy & Technology* 33 (1): 9-31.

Carnap, Rudolf. 1956. Empiricism, Semantics and Ontology. In *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: The University of Chicago Press.

Chalmers, David J. 1995. Facing up to the Problem of Consciousness. *Journal of Consciousness Studies* 2 (3): 200-19.

Chappell, Timothy. 2011. On the Very Idea of Criteria for Personhood. *The Southern Journal of Philosophy*. 49 (1): 1-27.

Coeckelbergh, Mark. 2010. Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology* 12: 209–21.

———. 2012. *Growing Moral Relations: Critique of Moral Status Ascription.* New York: Palgrave Macmillan.

———. 2014. The Moral Standing of Machines: Towards a Relational and Non-Cartesian Moral Hermeneutics. *Philosophy & Technology* 27 (1): 61-77.

———. 2018. Why Care About Robots? Empathy, Moral Standing, and the Language of Suffering. *Kairos. Journal of Philosophy & Science* 20 (1): 141-58.

Coeckelbergh, Mark, and David J. Gunkel. 2014. Facing Animals: A Relational, Other-Oriented Approach to Moral Standing. *Journal of Agricultural and Environmental Ethics* 27: 715-33.

Copp, David, ed. 2005. *The Oxford Handbook of Ethical Theory.* Oxford: Oxford University Press.

Cozzo, Cesare. 2012. Gulliver, Truth and Virtue. *Topoi: An International Review of Philosophy* 31(1): 59-66.

Craig, Edward. 2007. Genealogies and the State of Nature. In *Bernard Williams*, ed. Alan Thomas. Cambridge: Cambridge University Press.

———. 1990. *Knowledge and the State of Nature: An Essay In Conceptual Synthesis.* Oxford: Oxford University Press.

Crary, Alice. 2010. Minding What Already Matters: A Critique of Moral Individualism. *Philosophical Topics* 38 (1): 17-49.

Danaher, John. 2017. Should We Be Thinking About Sex Robots? In *Robot Sex: Social Implications and Ethical*, ed. John Danaher and Neil McArthur. Cambridge: MIT Press.

———. 2019. The Rise of the Robots and the Crisis of Moral Patiency. *AI & Society* 34 (1): 129–36

———. 2020. Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism. *Science and Engineering Ethics* 26 (4): 2023–49.

———. 2019. The Philosophical Case for Robot Friendship. *Journal of Posthuman Studies* 3 (1): 5-24. https://doi.org/10.5325/jpoststud.3.1.0005

Darling, Kate. 2016. Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects. In *Robot Law*, ed. Ryan Calo, A. Michael Froomkin, and Ian Kerr. Northampton: Edward Elgar.

———. 2021. *The New Breed: What Our History with Animals Reveals about Our Future with Robots*. New York: Henry Holt & Company.

Davidson, Donald. 1978/2000. What Metaphors Mean. In *Perspectives in the Philosophy of Language*, ed. Robert J. Stainton. Peterborough: Broadview Press.

———. 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

De Caro, Mario, and David Macarthur. 2004. *Naturalism in Question*. Cambridge: Harvard University Press.

———. 2010. *Naturalism and Normativity*. New York: Columbia University Press.

DeGrazia, David. 1996. *Taking Animals Seriously: Mental Life and Moral Status*. Cambridge: Cambridge University Press.

———. 2007. Human-Animal Chimeras: Human Dignity, Moral Status, and the Species Prejudice. *Metaphilosophy* 38 (2-3): 309-29.

———. 2008. Moral Status As a Matter of Degree? *Southern Journal of Philosophy* 46 (2): 181-98.

———. 2016. Modal Personhood and Moral Status: A Reply to Kagan's Proposal. *Journal of Applied Philosophy* 33 (1): 22-25.

———. 2019. Human-Animal Chimeras, “Human” Cognitive Capacities, and Moral Status. Hastings Center Report 49 (5): 33-34.

Dewey, John. 1922/1983. *Human Nature and Conduct: The Middle Works, 1899-1924, Volume 14: 1922,* ed. Jo Ann Boydston. Carbondale: Southern Illinois University Press.

———. 1925/2010. *Experience and Nature*. Whitefish: Kessinger Legacy Reprints.

———. 1929/2004. *Reconstruction in Philosophy*. Mineola, NY: Dover Publications.

———. 1929/1984. *The Quest for Certainty: The Later Works, 1925-1953, Volume 4: 1929*, ed. Jo Ann Boydston. Carbondale: Southern Illinois University Press.

Diamond, Cora. 1978. Eating Meat and Eating People. *Philosophy* 53 (206): 465-79.

———. Bernard Williams on the Human Prejudice. *Philosophical Investigations* 41 (4): 379-98

Dieleman, Susan. 2012. Solving the Problem of Epistemic Exclusion: A Pragmatist Feminist Approach. In *Contemporary Feminist Pragmatism*, ed. Maurice Hamington and Celia Bardwell-Jones. New York: Routledge.

DiSilvestro, Russell. 2010. *Human Capacities and Moral Status*. New York: Springer.

———. 2006. Not Every Cell Is Sacred: A Reply to Charo. *Bioethics* 20 (3): 146-57.

Dombrowski, Daniel A. 1984. Vegetarianism and the Argument from Marginal Cases in Porphyry. *Journal of the History of Ideas* 45 (1): 141-43.

Douglas, Thomas. 2013. Human Enhancement and Supra-personal Moral Status. *Philosophical Studies* 162 (3): 473-97.

Dreier, J. 2004. Meta-ethics and the Problem of Creeping Minimalism. *Philosophical Perspectives* 18 (1): 23-44.

Dutilh Novaes, Catarina. 2015. Conceptual Genealogy For Analytic Philosophy. In *Beyond the Analytic-Continental Divide: Pluralist Philosophy in the Twenty-First Century*, ed. Jeffery A. Bell, Andrew Culrofello, and Paul M. Livingston. London: Routledge.

Floridi, Luciano. 1999. Information ethics: On the Philosophical Foundation of Computer Ethics. *Ethics and Information Technology* 1: 33–52

———. 2002. On the Intrinsic Value of Information Objects and the Infosphere. *Ethics and Information Technology* 4: 287–304.

Ford, Martin. 2015. *Rise of the Robots: Technology and the Threat of a Jobless Future*. New York: Basic Books.

Frank, Lily, and Sven Nyholm. 2017. Robot sex and consent: Is consent to sex between a robot and a human conceivable, possible, and desirable? *Artificial Intelligence and Law* 25: 305–23.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

———. 2016. What's the Point of Blame? A Paradigm Based Explanation. *Noûs* 50 (1): 165-83.

Gibbard, Allan. 1992. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambrdge: Harvard University Press.

Gill, Michael B. 2006. *The British Moralists on Human Nature and the Birth of Secular Ethics*. Cambridge: Cambridge University Press.

Gleeson. 2008. Eating Meat and Reading Diamond. *Philosophical Papers* 37 (1): 157-75.

Gordon, John-Stewart. 2021. Artificial Moral and Legal Personhood. *AI & Society* 36 (2): 457–71.

Gross, Steven, Nicholas Tebben, and Michael Williams. 2015. *Meaning Without Representation: Essays on Truth, Expression, Normativity, and Naturalism.* Oxford: Oxford University Press.

Gruen, Lori, The Moral Status of Animals, *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/moral-animal/>.

Guignon, Charles, and David R. Hiley. 2003. Introduction: Richard Rorty and Contemporary Philosophy. In *Richard Rorty*, ed. Charles Guignon and David R. Hiley. Cambridge: Cambridge University Press.

Gunkel, David J. 2011. *The Machine Question*. Cambridge: MIT Press.

———. 2014. A Vindication of the Rights of Machines. *Philosophy and Technology* 27 (1):113–32.

———. 2018. The Other Question: Can and Should Robots Have Rights? *Ethics and Information Technology* 20: 87–99.

Haack, Susan. 1988. Surprising Noises: Rorty and Hesse on Metaphor. *Proceedings of the Aristotelian Society* 88 (1): 293-301.

Harman, Elizabeth. 2007. Sacred Mountains and Beloved Fetuses: Can Loving or Worshipping Something Give It Moral Status? *Philosophical Studies* 133 (1): 55-81.

———. 2003. The Potentiality Problem. *Philosophical Studies* 114 (1): 173–98.

Harvey, Graham. 2014. *The Handbook of Contemporary Animism*. New York: Routledge.

Heney, Diana B. 2015. Reality as Necessary Friction. *The Journal of Philosophy* 112 (9): 504–14.

———. 2016. *Toward a Pragmatist Metaethics*. New York: Routledge.

Hobbes, Thomas. 1996/1651. *Leviathan*. Cambridge: Cambridge University Press.

Horta, Oscar. 2017. Why the Concept of Moral Status Should Be Abandoned. *Ethical Theory and Moral Practice* 20 (4): 899–910.

Horwich, Paul. 1998. *Truth* (second edition). New York: Oxford University Press.

———. 2013. Naturalism, Deflationism, and the Relative Priority of Language and Metaphysics. In *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press.

Hume, David. 1978/1740. *A Treatise of Human Nature*. Oxford: Oxford University Press.

Hymers, Michael. 1998. Metaphor, Cognitivity, and Meaning-Holism. *Philosophy & Rhetoric* 31 (4): 266-82.

Jackman, Henry. 2019. William James on Moral Philosophy and its Regulative Ideals. *William James Studies* 15 (2): 1-25.

James, Williams. 1943/1907. *Pragmatism and Four Essays From The Meaning of Truth*. New York: Meridian Books.

———. 1956. *The Will to Believe: And Other Essays in Popular Philosophy*. New York: Dover Publications.

———. 1956/1891. The Moral Philosopher and the Moral Life. In *The Will to Believe: And Other Essays in Popular Philosophy*. New York: Dover Publications.

Jaworska, Agniezka. 2007. Caring and Full Moral Status. *Ethics* 117 (3): 460-97.

Jaworska, Agnieszka, and Julie Tannenbaum. 2014. Person-Rearing Relationships as a Key to Higher Moral Status. *Ethics* 124 (2): 242–71.

———. 2015. Who has the Capacity to Participate as a Rearee in a Person-Rearing Relationship? *Ethics* 125 (5): 1096-113.

———. 2018. The Grounds of Moral Status, *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/grounds-moral-status/>.

Jenkins, Richard. 2014. *Social Identity* (fourth edition). New York: Routledge.

Johnson, Mark. 2007. *The Meaning of the Body: Aesthetics of Human Understanding*. Chicago: University of Chicago Press.

———. 2014. *Morality for Humans: Ethical Understanding from the Perspective of Cognitive*

Science. Chicago: University of Chicago Press.

Kagan, Shelly. 2016. What's Wrong with Speciesism. *Journal of Applied Philosophy*. 33 (1): 1-21.

Kenny, Anthony. 1973. *Wittgenstein*. Cambridge: Harvard University Press.

Kitcher, Phillip. 2011. *The Ethical Project*. Cambridge: Harvard University Press

———. 2017. Social Progress. *Social Philosophy and Policy* 34 (2): 46-65.

Kittay, Eva Feder. 2005. At the Margins of Moral Personhood. *Ethics* 116 (1): 100–31.

Koplin, Julian J. 2019. Human-Animal Chimeras: The Moral Insignificance of Uniquely Human Capacities. *The Hastings Center Report* 49 (5): 23-32.

Knowles, Jonathan. 2017. Global Expressivism and the Flight from Metaphysics. *Synthese* 194 (12): 4781–97.

Koopman, Colin. 2013. *Genealogy as Critique: Foucault and the Problems of Modernity*. Bloomington: University of Indiana Press.

———. 2015. Two Uses of Michel Foucault in Political Theory: Concepts and Methods in Giorgio Agamben and Ian Hacking. *Constellations* 22 (4): 571-85.

———. 2016. Transforming the Self amidst the Challenges of Chance: William James on 'Our Undisciplinables.' *Diacritics* 44 (4): 40–65. https://doi.org/10.1353/dia.2016.0019

Koplin, Julian J. 2019. Human-Animal Chimeras: The Moral Insignificance of Uniquely Human Capacities. *Hastings Center Report* 49 (5): 23-32.

Korsgaard, Christine. 1996. *The Sources of Normativity.* Cambridge: Cambridge University Press

———. 2009. *Self-Constitution: Agency, Identity, and Integrity.* Oxford: Oxford University Press.

———. 2018. *Fellow Creatures: Our Obligations to the Other Animals*. Oxford: Oxford University Press.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kusch, Martin. 2009. Testimony and the Value of Knowledge. In *Epistemic Value*, ed. Adrian Haddock, Alan Millar, and Duncan Pritchard. Oxford: Oxford University Press.

———. 2013. Naturalized Epistemology and the Genealogy of Knowledge. In *Contemporary Perspectives on Early Modern Philosophy: Nature and Norms in Thought*, ed. Martin Lenz and Anik Waldow. Dordrecht: Springer.

Kusch, Martin, and Robin McKenna. 2020. The Genealogical Method in Epistemology. *Synthese* 197 (3): 1057-76.

Lakoff, George, and Mark Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

Legg, Catherine, and Paul Giladi. 2018. Metaphysics-Low in Price, High in Value: A Critique of Global Expressivism. *Transactions of the Charles S. Peirce Society: A Quarterly Journal in American Philosophy* 54 (1): 64–83.

Leiter, Brian. 2004. Introduction. In *The Future for Philosophy*, ed. Brian Leiter. Oxford: Oxford University Press.

Lekan, Todd. 2018. Who Are Moral Philosophers? Ethics William James Style. *The Pluralist*, 13 (1): 81–96.

Liao, S. Matthew. 2010. The Basis of Human Moral Status. *Journal of Moral Philosophy* 7 (2): 159-79.

Lindemann, Hilde. 2019. *An Invitation to Feminist Ethics Second Edition*. Oxford: Oxford University Press

Locke, John. 1988/1689. *Two Treatises of Government*. Ed. Peter Laslett. Cambridge: Cambridge University Press.

Lynch, Michael Patrick. 2004. *True to Life:Why Truth Matters*. Cambridge: The MIT Press.

———. 2015. Normativity, Pragmatism, and the Price of Truth. In *Meaning Without Representation : Essays on Truth, Expression, Normativity, and Naturalism* (first edition). Oxford: Oxford University Press.

Macarthur, David. 2008. Pragmatism, Metaphysical Quietism, and the Problem of Normativity. *Philosophical Topics* 36 (1): 193-209.

———. 2014. Subject Naturalism, Scientism and the Problem of Linguistic Meaning: Critical Remarks on Price's "Naturalism Wtihout Representationalism." *Análisis: Revista de Investigación Filosfófica* 1 (1): 69-85.

———. 2015. Liberal Naturalism and Second-Personal Space: A Neo-Pragmatist Response to The Natural Origins of Content. *Philosophia* 43 (3): 565–78.

Macarthur, David, and Huw Price. 2011. Pragmatism, Quasi-realism and the Global Challenge. In *Naturalism Without Mirrors*. Oxford: Oxford University Press.

Marchetti, Sarin. 2015. *Ethics and Philosophical Critique in William James.* London: Palgrave Macmillan.

Marder, Michael. 2013. *Plant-Thinking: A Philosophy of Vegetal Life.* New York: Columbia University Press.

Marti, Patrizia. 2010. Robot Companions. *Interaction Studies* 11 (2): 220-26.

McMahan, Jeff. 2005. Our Fellow Creatures. *The Journal of Ethics* 9 (3/4): 353–80.

———. 2002. *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press.

Miller, Alexander. 2003. *An Introduction to Contemporary Metaethics*. Cambridge: Polity.

Miller, Daniel. 1994. Artefacts and the Meaning of Things. In *Companion Encyclopedia of Anthropology* (second edition), ed. Tim Ingold. London: Routledge.

Misak, Cheryl. 2015. Pragmatism and the Function of Truth. In *Meaning Without Representation: Essays on Truth, Expression, Normativity, and Naturalism* (first edition), ed. Steven Gross, Nicholas Tebben, Michael Williams. Oxford: Oxford University Press.

Moody-Adams, Michele. 2017. Moral Progress and Human Agency. *Ethical Theory and Moral Practice* 20 (1): 153-68.

Moore, A. W. 1991. Can Reflection Destroy Knowledge? *Ratio* 4 (2): 97-106.

——— . 2003. Williams on Ethics, Knowledge, and Reflection. *Philosophy* 78 (305): 337-54.

Moore, G. E. 1962/1903. *Principia Ethica*. Cambridge: Cambridge University Press.

Mosakas, Kestutis. 2021. On the Moral Status of Social Robots: Considering the Consciousness Criterion. *AI & Society* 36 (3): 429–43.

Müller, Vincent C. 2021. Is it Time for Robot Rights? Moral Status in Artificial Entities. *Ethics and Information Technology*. https://doi.org/10.1007/s10676-021-09596-w

Mullhall, Stephen. 2002. Fearful Thoughts. *London Review of Books* 24 (16).

Musschenga, Albert W., and Gerben Meynen. 2017. Moral Progress: an Introduction. *Ethical Theory and Moral Practice* 20 (1): 3-15.

Neely, Erica L. 2014. Machines and the Moral Community. *Philosophy and Technology* 27 (1): 97-111.

Nguyen, C. Thi. 2021. How Twitter Gamifies Communication. In *Applied Epistemology*, ed. Jennifer Lackey. Oxford: Oxford University Press.

Nietzsche, Friedrich. 2006/1887. *Nietzsche: "On the Genealogy of Morality" and Other Writings*, ed. Keith Ansell-Pearson. Trans. Carol Diethe. Cambridge: Cambridge University Press.

Nørskov, Marco. 2016. *Social Robots: Boundaries, Potential, Challenges*. London: Routledge.

Parviainen, Jaana, and Mark Coeckelbergh. 2021. The Political Choreography of The Sophia Robot: Beyond Robot Rights and Citizenship to Political Performances for the Social Robotics Market. *AI & Society* 36 (2): 715–24.

Popper, Karl. 1963. *Conjectures and Refutations: The Growth of Scientific Knowledge.* New York: Basic Books.

Price, Huw. 1988. *Facts and the Function of Truth: Philosophy of Theory*. Cambridge: Cambridge University Press.

———. 1992. Metaphysical Pluralism. In *Naturalism Without Mirrors* (2011). Oxford: Oxford University Press.

——— . 1993. Semantic Minimalism and the Frege Point. In *Naturalism Without Mirrors* (2011). Oxford: Oxford University Press.

———. 1996. Review of Simon Blackburn, Essays in Quasi-realism. In *Philosophy and Phenomenological Research* 56 (4): 965—68.

———. 1997. What Should a Deflationist About Truth Say About Meaning? *Philosophical Issue*s 8: 107–15.

———. 1998. Three Norms of Assertibility, or How the Moa Became Extinct. *Philosophical Perspectives* 12: 241-54.

———. 2003. Truth as Convenient Friction. In *Naturalism Without Mirrors* (2011). Oxford: Oxford University Press.

———. 2004. Immodesty Without Mirrors: Making Sense of Wittgenstein's Linguistic Pluralism. In *Naturalism Without Mirrors* (2011). Oxford: Oxford University Press.

———. 2006. Blackburn and the War on Error. A Discussion Review of Simon Blackburn's Truth: A Guide for the Perplexed. In *Australasian Journal of Philosophy* 84: 603—614.

———. 2009a. Metaphysics After Carnap: The Ghost Who Walks? In *Naturalism Without Mirrors* (2011). Oxford: Oxford University Press.

———. 2009b. The Semantic Foundations of Metaphysics. In *Naturalism Without Mirrors* (2011). Oxford: Oxford University Press.

———. 2010. One Cheer for Representationalism? In *The Philosophy of Richard Rorty*, ed. Randall E. Auxier and Lewis Edwin Hahn. Chicago: Open Court.

———. 2011. *Naturalism Without Mirrors*. Oxford: Oxford University Press.

———. 2013. *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press.

Queloz, Mathieu. 2018a. How Genealogies Can Affect the Space of Reasons. *Synthese*: 1-23.

———. 2018b. Williams's Pragmatic Genealogy and Self-Effacing Functionality. *Philosopher's Imprint* 18 (17): 1-20.

———. 2019. From Paradigm-Based Explanation to Pragmatic Genealogy. *Mind* 129 (515): 683-714.

Queloz, Matthieu. 2021. *The Practical Origins of Ideas: Genealogy as Conceptual Reverse-Engineering*. Oxford: Oxford University Press

Rachels, James. 2005. Drawing Lines. In *Animal Rights: Current Debates and New Directions*, ed. Cass R. Sunstein and Martha C. Nussbaum. Oxford: Oxford University Press.

Railton, Peter. 1986. Moral Realism. *The Philosophical Review* 95 (2): 163–207.

Redding, Paul. 2010. Two Directions for Analytic Kantianism: Naturalism and Idealism in *Naturalism and Normativity*, ed. Mario De Caro and David MacArthur. New York: Columbia University Press.

Regan, Tom. 1983. *The Case for Animal Rights*. Berkeley: University of California Press.

———. 1986/87. The Case for Animal Rights. In Advances in Animal Welfare Science, ed. M.W. Fox and L.D. Mickley. Washington, D.C.: *The Humane Society of the United States*.

Robert, Jason Scott, and Françoise Baylis. 2003. Crossing Species Boundaries. *American Journal of Bioethics* 3 (3): 1-13.

Roojen, Mark Van. 1996. Expressivism and Irrationality. *The Philosophical Review* 105 (3): 311–35.

Rorty, Richard. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.

———. 1989. *Continency, Irony, and Solidarity.* Cambridge: Cambridge University Press.

———. 1991a. Unfamiliar Noises: Hesse and Davidson on Metaphor. In *Objectivity, Relativism, and Truth: Philosophical Papers Volume 1*. Cambridge: Cambridge University Press.

———. 1991b. Philosophy as Science, as Metaphor, and as Politics. In *Essays on Heidegger and Others*: *Philosophical Papers Volume 2.* Cambridge: Cambridge University Press.

———. 1998. Is Truth the Goal of Inquiry? Donald Davidson versus Crispin Wright. In *Truth and Progress: Philosophical Papers Volume 3*. Cambridge: Cambridge University Press.

———. 2002. To the Sunlit Uplands (Review of *Truth and Truthfulness*). *London Review of Books* 24 (21).

———. 2007. *Philosophy as Cultural Politics: Philosophical Papers Volume 4*. Cambridge: Cambridge University Press.

Rorty, Richard, and Huw Price. 2010. Exchange On Truth As Convenient Friction. In *Naturalism and Normativity*, ed. Mario De Caro and David Macarthur. New York: Columbia University Press.

Rosen, Gideon. 1998. Blackburn's Essays in Quasi-Realism. *Nous* 32 (3): 386–405.

Rousseau, Jean-Jacques. 1987/1782. *The Basic Political Writings*. Trans. Peter Gay. Indianapolis: Hackett Publishing Company.

Rydenfelt, Henrik. 2019. Realism without Representationalism. *Synthese* 198: 2901-2918. https://doi.org/10.1007/s11229-019-02251-4

Sachs, Benjamin. 2011. The Status of Moral Status. *Pacific Philosophical Quarterly* 92 (1): 87–104.

Sagoff, Mark. 1984. Animal Liberation and Environmental Ethics: Bad Marriage, Quick Divorce. *Osgoode Hall Law Journal* 22 (2): 297-307.

Savulescu, Julian. 2009. The Human Prejudice and the Moral Status of Enhanced Beings: What Do We Owe the Gods? In *Human Enhancement*, ed. Nick Bostrom and Julian Savulescu. Oxford: Oxford University Press.

Scanlon, T. M. 1998. *What We Owe to Each Other.* Cambridge: Harvard University Press.

Schechtman, Marya. 2014. *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. Oxford: Oxford University Press.

Schinkel, Anders, and Doret de Ruyter. 2017. Individual Moral Development and Moral Progress. *Ethical Theory and Moral Practice* 20 (1): 121–36.

Schneider, Susan. 2019. *Artificial You: AI and the Future of Your Mind*. Princeton: Princeton University Press.

Schwitzgebel, Eric, and Mara Garza. 2015. A Defense of the Rights of Artificial Intelligences. *Midwest Studies In Philosophy* 39 (1): 98-119.

Sharkey, Amanda, and Noel Sharkey. 2012. Granny and the Robots: Ethical Issues in Robot Care for the Elderly. *Ethics and Information* 14: 27–40.

Shepard, Joshua. 2018. *Consciousness and Moral Status.* New York: Routledge.

Sellars, Wilfrid. 1968. *Science and Metaphysics: Variations on Kantian Themes*. New York: Routledge & K. Paul.

———. 1997/1956. *Empiricism and the Philosophy of Mind*. Cambridge: Harvard University Press.

Shapiro, Lionel. 2014. Linguistic Function and Content: Reflections on Price's Pragmatism. *The Philosophical Quarterly* 64 (256): 497-506.

Shoemaker, David W. 2007. Personal Identity and Practical Concerns. *Mind* 116 (462): 317–57.

Silvers, Anita. 2012. Moral Status: What a Bad Idea! *Journal of Intellectual Disability Research* 56 (11): 1014–25.

Singer, Peter. 1974. All Animals Are Equal. *Philosophical Exchange* 1 (5): 103-16.

———. 2009. Speciesism and Moral Status. *Metaphilosophy* 40 (3–4): 567–81.

———. 2011. *Practical Ethics* (third edition). Cambridge: Cambridge University Press.

Slater, Michael R. 2007. Ethical Naturalism and Religious Belief in 'The Moral Philosopher and the Moral Life'. *William James Studies* 2 (1).

Smyth, Nicholas. 2017. The Function of Morality. *Philosophical Studies* 174 (5): 1127-44.

Srinivasan, Amia. 2019. Genealogy, Epistemology, and Worldmaking. *Proceedings of the Aristotelian Society,* Vol. CXIX: 127-56.

Steiner, Gary. 2005. *Anthropocentrism and Its Discontents*. Pittsburgh: University of Pittsburgh.

Street, Sharon. 2012. Coming to Terms with Contingency: Humean Constructivism About Practical Reason. In *Constructivism in Practical Philosophy*, ed. Jimmy Lenman and Yonatan Shemmer. Oxford: Oxford University Press.

———. 2006. A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*. 127 (1): 109–66.

Streiffer, Robert. 2005. At the Edge of Humanity: Human Stem Cells, Chimeras, and Moral Status. *Kennedy Institute of Ethics Journal* 15 (4): 347-70.

———. 2019. Human/Non-Human Chimeras, *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2019/entries/chimeras/>.

Tavani, Herman T. 2018. Can Social Robots Qualify for Moral Consideration? Reframing the Question About Robot Rights. *Information* 9 (4), 73.

Tebben, Nicholas. 2015. Introduction. In *Meaning Without Representation : Essays on Truth, Expression, Normativity, and Naturalism* (first edition). Oxford: Oxford University Press.

Thomson, Judith Jarvis. 1975. The Right to Privacy. *Philosophy & Public Affairs* 4 (4): 295-314.

Warren, Mary Anne. 1997. *Moral Status: Obligations to Persons and Other Living Things*. Oxford: Oxford University Press.

Wasserman, David, Adrienne Asch, Jeffrey Blustein, and Daniel Putnam. 2017. Cognitive Disability and Moral Status, *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/cognitive-disability/>.

Williams, Bernard. 1986. *Ethics and the Limits of Philosophy*. Cambridge: Harvard University Press.

———. 2002. *Truth and Truthfulness*. Princeton: Princeton University Press.

———. 2005. *In the Beginning Was the Deed: Realism and Moralism in Political Argument*. Princeton: Princeton University Press.

———. 2006. *Philosophy as a Humanistic Discipline*. Princeton: Princeton University Press.

Williams, Michael. 2013. How Pragmatists Can Be Local Expressivists. In *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press.

Williamson, Timothy. 2004. Past the Linguistic Turn? In *The Future for Philosophy*, ed. Brian Leiter. Oxford: Oxford University Press.

Wittgenstein, Ludwig. 1953/2009. *Philosophical Investigations* (revised fourth edition), ed. P. M. S. Hacker and Joachim Schulte. Trans. G. E. M. Anscombe, P. M. S. Hacker and Joachim Schulte. Malden: Wiley-Blackwell.

Wray, K. Brad. 2011. *Kuhn's Evolutionary Social Epistemology*. Cambridge: Cambridge University Press.

Wrenn, Corey Lee. 2019. The Vegan Society and Social Movement Professionalization, 1944–2017. *Food and Foodways: Explorations in the History and Culture of Human Nourishment* 27 (3): 190–210.