CONSCIENCE, HUMAN NATURE, AND THE EVOLUTIONARY CHALLENGE

by

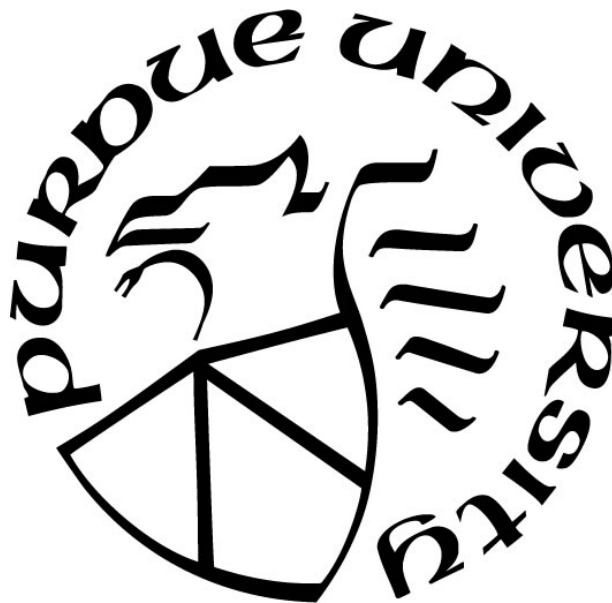**Brian M. Johnson**

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Philosophy

West Lafayette, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Christopher Yeomans, Chair

    Chair of the Department of Philosophy, College of Liberal Arts

Dr. Patrick Kain

    Department of Philosophy, College of Liberal Arts

Dr. Daniel R. Kelly

    Department of Philosophy, College of Liberal Arts

Dr. Daniel Smith

    Department of Philosophy, College of Liberal Arts

Approved by:

    Dr. Christopher Yeomans

        Head of the Graduate Program

For Patricia Johnson

ACKNOWLEDGMENTS

TABLE OF CONTENTS

ABBREVIATIONS

Works by Georg Wilhelm Friedrich Hegel

Enz   *Enzyklopädie der Philosophischen Wissenschaften im Grundrisse* in *Werke in zwanzig Bänden.* Edited by Eva Moldenhauer and Karl Markus Michel. Frankfurt am Main: Suhrkamp Verlag, 1986. Cited by paragraph (p.) number.

ETW   *Early Theological Writings,* translated by T.M. Knox. Philadelphia: University of Pennsylvania Press, 1975.

GPR   *Grundlinien der Philosophie des Rechts.* Fourth Edition. Edited by Johannes Hoffmeister. Berlin: Akademie Verlag, 1955. English quotations are from *Elements of the Philosophy of Right.* Edited by Allen Wood and Translated by H.B. Nisbet. Cambridge: Cambridge University Press, 1991.

PdG   *Phänomenologie des Geistes.* Edited by Johannes Hoffmeister. Hamburg: Felix Mainer Verlag, 1975. English quotations are from *Phenomenology of Spirit.* Translated by A.V. Miller. Oxford: Oxford University Press according to Miller's paragraph (p.) numbers.

WTA   "Who Thinks Abstractly," in *Hegel: A Reinterpretation, Texts, and Commentary.* Ed. Walter Kaufmann. New York: Doubleday, 1965.

Works by Immanuel Kant

EMH   "Essay on the Maladies of the Head." References are by volume and page number to Kant's *Gesammelte Werke.* Edited by the Deutsche Akademie der Wissenschaften. Berlin: De Gruyter, 1908. English quotations are taken from *Anthropology, History, and Education.* Edited by Günter Zöller and Robert Louden. Cambridge: Cambridge University Press, 2007.

G   *Grundlegung zur Metaphysic der Sitten.* References are by volume and page number to Kant's *Gesammelte Werke.* Edited by the Deutsche Akademie der Wissenschaften. Berlin: De Gruyter, 1908. English quotations are taken from *Practical Philosophy.* References are by volume and page to the Edited by M. Gregor. Cambridge: Cambridge University Press, 1996.

IUH   "Idea for a Universal History with Cosmopolitan Intent." References are by volume and page number to Kant's *Gesammelte Werke.* Edited by the Deutsche Akademie der Wissenschaften. Berlin: De Gruyter, 1908. English

quotations are taken from *Anthropology, History, and Education*. Edited by Günter Zöller and Robert Louden. Cambridge: Cambridge University Press, 2007.

KpV   *Kritik der praktischen Vernunft.* References are by volume and page number to Kant's *Gesammelte Werke.* Edited by the Deutsche Akademie der Wissenschaften. Berlin: De Gruyter, 1908. English quotations are taken from *Practical Philosophy.* Edited by M. Gregor. Cambridge: Cambridge University Press, 1996.

LP   *Lectures on Pedagogy,* References are by volume and page number to Kant's *Gesammelte Werke.* Edited by the Deutsche Akademie der Wissenschaften. Berlin: De Gruyter, 1908. English quotations are taken from *Anthropology, History, and Education.* Edited by Günter Zöller and Robert Louden. Cambridge: Cambridge University Press, 2007.

MdS   *Metaphysik der* Sitten. References are by volume and page number to Kant's *Gesammelte Werke.* Edited by the Deutsche Akademie der Wissenschaften. Berlin: De Gruyter, 1908. English quotations are taken from *Practical Philosophy.* Edited by M. Gregor. Cambridge: Cambridge University Press, 1996.

OBS   *Observations on the Feeling of the Beautiful and the Sublime.* References are by volume and page number to Kant's *Gesammelte Werke.* Edited by the Deutsche Akademie der Wissenschaften. Berlin: De Gruyter, 1908. English quotations are taken from *Anthropology, History, and Education.* Edited by Günter Zöller and Robert Louden. Cambridge: Cambridge University Press, 2007.

PP   "Perpetual Peace.*"* References are by volume and page number to Kant's *Gesammelte Werke.* Edited by the Deutsche Akademie der Wissenschaften. Berlin: De Gruyter, 1908. English quotations are taken from *Practical Philosophy.* Edited by M. Gregor. Cambridge: Cambridge University Press, 1996.

ODC   "On the Determination of the Concept of Race." References are by volume and page number to Kant's *Gesammelte Werke.* Edited by the Deutsche Akademie der Wissenschaften. Berlin: De Gruyter, 1908. English quotations are taken from *Anthropology, History, and Education.* Edited by Günter Zöller and Robert Louden. Cambridge: Cambridge University Press, 2007.

ODR   "Of the Different Races of Human Beings" References are by volume and page number to Kant's *Gesammelte Werke.* Edited by the Deutsche Akademie der Wissenschaften. Berlin: De Gruyter, 1908. English quotations are taken from *Anthropology, History, and Education.* Edited

by Günter Zöller and Robert Louden. Cambridge: Cambridge University Press, 2007.

OTP   "On the Use of Teleological Principles in Philosophy." References are by volume and page number to Kant's *Gesammelte Werke*. Edited by the Deutsche Akademie der Wissenschaften. Berlin: De Gruyter, 1908. English quotations are taken from *Anthropology, History, and Education*. Edited by Günter Zöller and Robert Louden. Cambridge: Cambridge University Press, 2007.

WHD   "What does it Mean to Orient Oneself in Thinking," References are by volume and page number to Kant's *Gesammelte Werke*. Edited by the Deutsche Akademie der Wissenschaften. Berlin: De Gruyter, 1908. English quotations are taken from *Anthropology, History, and Education*. Edited by Günter Zöller and Robert Louden. Cambridge: Cambridge University Press, 2007.

Works by Jean-Jacques Rousseau

OC   *Oeuvres Complètes*. Edited by Bernard Gagnebin and Marcel Raymond. Paris: Pléiade, 1959.

EE   *Emile, ou de l'Education*. English quotations are taken from *Emile, or On Education*. Translated by Allen Bloom. New York: Basic Books, 1979.

CGP   *Considerations on the Government in Poland*. References are to the English translation in *The Social Contract and Other Later Political Writings*. Edited and translated by Victor Gourevitch. Cambridge: Cambridge University Press, 1997.

DOI   *Discourse on the Origin and Foundations of Inequality*. References are to the English translation in *The* Discourses *and Other Early Political Writings*. Edited and Translated by Victor Gourevitch. Cambridge: Cambridge University Press, 1997.

DPE   *Discourse on Political Economy*. References are to the English translation in *The Social Contract and Other Later Political Writings*. Edited and translated by Victor Gourevitch. Cambridge: Cambridge University Press, 1997.

SC   *The Social Contract*. References are to the English translation in *The Social Contract and Other Later Political Writings*. Edited and translated by Victor Gourevitch. Cambridge: Cambridge University Press, 1997.

ABSTRACT

Author: Johnson, Brian M. PhD
Institution: Purdue University
Degree Received: May 2019
Title: Conscience, Human Nature, and the Evolutionary Challenge
Committee Chair: Dr. Christopher Yeomans

The purpose of this dissertation is to rebut some skeptical arguments in moral epistemology by appealing to philosophical resources from the history of European philosophy. The skeptical arguments I will be countering are grounded in the perspective of contemporary biology. Put quickly, our evolutionary history is said to undermine our claims to moral knowledge because the process by which our capacity for such knowledge developed was determined by adaptive and reproductive fitness. The determinations of fitness, it is said, cannot be expected to align with standards of objective moral value. In the first chapter, I spell out the importance of evaluative perception. The need for a capacity to perceive value raises the concern that moral psychology is something mysterious. In the second chapter, I consider some skeptical arguments in moral epistemology that conclude we have no good reason to believe we are wired to be receptive to objective moral truth. While some of these arguments purport to undermine our access to objective moral truths, I conclude that they do not. The remainder of the dissertation considers the work of Rousseau, Kant, and Hegel in light of the concerns raised in part one. Despite initial appearances, these authors understand the relation of conscience and human nature in a way that points toward a defensible view, even in light of the challenges raised by contemporary biology. The resulting view is an account of the moral conscience that emphasizes

autonomy and rational agency and recognizes their value in virtue of their concrete expression in a social context.

INTRODUCTION

On the History of Philosophy and its Methods

In this dissertation I will attempt to rebut some skeptical arguments in moral epistemology by appealing to philosophical resources in the work of some late-eighteenth and early-nineteenth century European philosophers. The skeptical arguments I will be countering are grounded in the perspective of contemporary biology. Put quickly, our evolutionary history is said to undermine our claims to moral knowledge because the process by which our capacity for such knowledge developed was determined by adaptive and reproductive fitness, which cannot be expected to align with standards of objective moral value.

In the first chapter, I spell out the importance of what I call 'evaluative perception.' The need for a capacity to perceive value raises the question of whether we are indeed wired for such perception. In the second chapter, I consider some arguments that conclude we have no good reason to believe that we are wired to be receptive to objective moral truth. While some of these arguments purport to undermine our ability to perceive objective value, I conclude that they do not. The remainder of the dissertation considers what Rousseau, Kant, and Hegel have to say about the relation of the moral conscience to human nature and how well these views hold up to the skeptical arguments discussed in the first part of the dissertation.

The project of appealing to pre-Darwinian figures to confront arguments premised on contemporary biology naturally invites concerns about anachronism. I will attempt to briefly address these concerns by saying a little bit about the history of philosophy and debates

surrounding methodology. To illustrate the point, I will briefly touch on an issue that has gotten a great deal of attention in recent decades, Kant's racism.

First off, it's not immediately obvious that the fact that Rousseau came before Darwin should be a problem. Obviously, there is no way that Rousseau could have responded directly to the implications of biological discoveries that occurred long after he died. But it doesn't follow that such findings cause trouble for his view. The arguments I am addressing are ethical. We cannot determine whether evolutionary debunking arguments undermine Rousseauian ethical philosophy until we have carefully laid out what Rousseau's position was. Only then can we consider the implications of the arguments themselves. If it is anachronistic hero-worshipping to obsess over what Rousseau *would* have said had he lived to see the success of evolutionary biology, it is no less arrogant to suppose from the outset that ideas from the past are of no use to us today. Evolutionary biology has profoundly improved our understanding of our species and the world in which we live. But to those who would take this as a reason to be dismissive of the history of philosophy, part of my goal here is to say, "Not so fast." Darwinian debunking arguments aren't as destructive to the core ethical theories of past philosophers as one might assume.

There are two well recognized approaches to the history of philosophy. Proponents of *rational reconstruction* occasionally express frustration at those philosophers they see as engaged in mere *intellectual history*. Philosophers of the past were trying to get at the truth, and it goes against the spirit of their enterprise to merely catalogue the history of their errors. [1] From this perspective, the history of philosophy is interesting to the extent that it gets things right, and

---

[1] Wood, 2008.

it is most interesting when a philosopher is seen to have made progress upon which a contemporary theory can build. At its best, rational reconstruction can use contemporary insights to show the presuppositions and inconsistencies that the original author may have missed. For example, Christine Korsgaard's essay on Kant's universal law formulation identifies three different interpretations of the term contradiction and defends her own practical contradiction interpretation.[2] In so doing, she identifies a characteristic inconsistency in Kant's own writing. And by developing her own practical interpretation, she makes a valuable contribution to contemporary normative ethics. Scholars who reject this approach see it as anachronism and distortion.[3] On this view, essays like Korsgaard's reduce 'Kant' to a name brand for a position in contemporary analytic philosophy.[4]

I am of the opinion that *intellectual history* and *rational reconstruction* both have their place in the history of philosophy. Both can be genuinely historical, and both can be genuinely philosophical. Both should be criticized when they misrepresent a philosopher's view. And both should be criticized when they prop-up bad ideas on the authority of a famous name. The difference is one of focus. Rational reconstruction puts the truth-question first. Focusing on what someone should have said may make it hard to notice when someone *just got it wrong*. Intellectual history puts the historical question first. This risks passing over subtle flashes of genuine philosophical insight and reducing a profound thinker to *a mere child of the age*.

---

[2] Korsgaard, 1985.

[3] Beiser, 1995, 2007.

[4] Bernasconi, 2003: 16.

Take Kant's racism for example. Kant is largely recognized as the great thinker of human dignity, autonomy, and equality. But while most scholars today think Kant's moral theory entails respect and dignity for all persons regardless of merely physiological differences, this wasn't obvious to Kant himself. Yet his views on race were grotesque.[5] To borrow a formulation from Charles Mills, we are left with the following choice: either Kant was an inconsistent egalitarian or a consistent inegalitarian. Mills adopts the latter view, arguing that Kant viewed the non-white races as less than human.[6] Others, adopting a more reconstructive approach, insist that racism is inconsistent with the critical core of the Kantian moral philosophy. Prejudice, according to this view, prevented Kant from fully grasping the implications of his own view.[7]

But this dichotomy between rational reconstruction and intellectual history unduly limits the ways in which the history of philosophy can be philosophical. Consider *genealogical critique*. There's an open philosophical question—which I will not be addressing here—concerning how autonomous philosophy can be. This question is implied by the suggestion that Kant was merely a product of his time. If this is true of Kant, it may also be true of us. If so, then the history of ideas becomes more important, since the way we view philosophical issues today depends heavily on their history. From the perspective of genealogical critique, it has been argued that Kant's views on race mark a pivotal moment in the dark history of the pseudo-science of biological racism.[8] If the ways of thinking available to us today are a function of the history of

---

[5] e.g. **ODR**, 2:427-43; **ODC**, 8:89-106; **OTP**, 8:174; and **OBS**, 2:243-55.

[6] Mills, 2005.

[7] Allais, 2016.

[8] Bernasconi, 2001.

ideas, we can learn something philosophically significant about ourselves by studying how it actually happened.

My approach in this dissertation is different. While these approaches all capture something important about the history of philosophy, they also miss something. There's something else which causes great philosophical texts to endure and remain relevant to new generations of readers. And these other approaches don't fit well with the way we often teach the history of philosophy to undergrads. Philosophers do philosophy. In so doing, they deal with perennial philosophical problems in sometimes unique and creative ways. In this dissertation, I will use the works as a touchstone for philosophical reflection on an important philosophical topic. These authors were smarter than I am, and I've learned a lot from them. I have tried not to project my contemporary sensibilities into their views. Instead, I've defended them on some fundamental metaethical points. I make no apologies for their sexism, racism, and other faults. Not because there's nothing to apologize for. There's much to apologize for; there's much to reject in their views. Nonetheless, many of the deepest and most fundamentally philosophical problems are perennial. Founding a theory of value involves stepping beyond the normal bounds of description, and when philosophers have attempted to account for *the good*, they have run into some characteristic difficulties. It may be that, by looking at how philosophers historically have engaged with the deepest problems of philosophy, we genuinely approach these themes ourselves —perhaps with new insights, perhaps in a new philosophical language. This way of approaching the history of philosophy may be, at once, more personal and more universal. But it's neither anachronistic nor unphilosophical.

If at least some philosophical questions are truly philosophical, we shouldn't be surprised that these get addressed in very different times and places. We also should not be surprised if such themes are addressed in radically different voices. One can go a long way toward appreciating how historical figures were determined by the age in which they wrote while also allowing that, sometimes, they were also determined by the deep structure of the object of their inquiry.

Some readers may be suspicious of the notion that there is any such 'deep structure' at work. These readers might call themselves constructivists and say that they reject realism. To such readers, I would make three quick points from the outset. (1) In my experience, realists and constructivists often talk past each other, and sometimes the main challenge in these debates is trying to figure out exactly what each side is even saying. (2) Whether there is anything real —'real' being such a sticky term—that philosophers are referring to when they talk about norms, there is certainly something real about us. Constructivists tend to acknowledge inherent limits to our ways of gaining access to whatever there may be 'out-there'—another less than optimal phrase. These limits may be enough to provide the kind of deep structure that I'm talking about. (3) Finally, in the discussion that follows, much will be said about perception. Here we must be cautious. Perception is easily conceived as a mode of access to the external world, but, as Kant's more sophisticated account of perception will help make clear, even seeing an empirical object involves systematically linking representations together according to a rule. If seeing an action as-*virtuous* turns out to fare as well as seeing a hunk of metal as-*a-spoon*, then things sound pretty good for the moralist.

Chapter Overview

My project has two goals. The first is to determine the implications of varieties of Darwinian debunking arguments in moral epistemology. The second is to develop a strategy for evading these arguments based on the accounts of conscience found in Rousseau, Kant, and Hegel.

Darwinian debunking arguments seek to undermine claims to moral knowledge by considering some facts about human evolution. They typically do this by considering the origin of human nature and then drawing inferences about the relation between our attitudes and moral truth. Consider the following from Sharon Street:

> It is clear, for instance, how fatal to reproductive success it would be to judge that the fact that something would endanger one's survival is a reason to do it, or the fact that someone is a kin is a reason to harm that individual. A creature who accepted such evaluative judgments would run itself off cliffs, seek out its predators, and assail its offspring, resulting in the speedy elimination of it and its evaluative tendencies from the world.[9]

Indeed, our ancestors would not have gotten far with a formula for judgment like that (if doing *x* will kill you, do *x*), but the very absurdity of such a formula is enough to show why a careful account of the actual origins of human nature is needed. Street's point is an general one, that *whatever* the correct evolutionary account of our moral beliefs turns out to be, it will entail that:

> tendencies to make certain evaluative judgments rather than others contributed to our ancestor's reproductive success not because they constituted perceptions of independent evaluative truths, but rather because they forged adaptive links between our ancestors' circumstances and their responses to those circumstances.[10]

What we have here is a very plausible general principle, that there is an 'adaptive link' between our ancestor's motivating attitudes and the conditions for survival in their environment. But the

---

[9] Street, 2006: 110.

[10] Street, 2006: 127.

details matter. How strong is this adaptive link? What kinds of behaviors are determined by it? Is there any room for free-play? If we suppose that underpinnings of human behavior involve multiple adaptive links to environment, what happens when the determinations of one functional adaptation conflict with another? It seems that if there is room for agents to act counter to these evolutionary determinations, then "perceptions of independent evaluative truths" could exist alongside our evolved responses and dispositions to judgment. Whatever the case may be, it doesn't seem likely that philosophers can resolve this issue from the armchair. In chapter two, I will argue that these Darwinian debunking arguments are at their best when they stick closely to what the science actually says and avoid speculations about what a completed science *would say.*

We can learn a great deal from accounts of the moral emotions that indicate ways in which an emotion is likely mislead us or grant outsized weight to certain evaluative judgments. But one thing is worth noting. As the story of the origins of moral feeling and moral belief becomes more detailed, it becomes increasingly evident that the influence of culture and education cannot be omitted from the genealogy of moral belief. Norm acquisition doesn't take place in a vacuum; it is sensitive to social cues.[11] These cues and the cultural context in which we encounter them needs to be included the story of the origin of our moral beliefs.

When we turn our attention to culture, the structure and justification of our values becomes more explicit. It turns out we can learn a great deal about what we value and why by considering the conscious efforts that human beings make in raising children. In a recent critique of neo-classical economic theory, Mariana Mazzucato argues that it is dishonest to suppose that

---

[11] For an excellent review of the literature on "self-domestication" in humans, see chapter eleven of Henrich, 2016.

economics can do without value theory, and she observes that educational policy gives the lie to the alleged neutrality of free-market capitalism:

> Deciding which activities are more important than others is critical in setting the direction for the economy: put simply, those activities thought to be more important in achieving particular objectives have to be increased and less important ones reduced. We already do this… We subsidize our education and training for students because as a society we want more young people to go to [college] or enter the workforce with better skills. Behind such policies may be economic models that show how investment in 'human capital' – people's knowledge and capabilities – benefits a country's growth by increasing productive capacity.[12]

That we subsidize education reflects an implicit valuation of what it is that education does. Fierce debates over education policy cannot show that it's about more than budgets. A commitment to universal education reflects a commitment to at least a certain degree of equality. Concerns about education undermining traditional values reflects a commitment to those values, and debates over how best to teach shameful aspects of our past reflect a commitment to coping with, and hopefully rectifying, past injustices. And even if eduction is primarily about promoting a particular kind or quality of material production, this still reflects a desire to see that production promoted. Education lays bare a society's implicit commitments by showing what we want for ourselves and for our children. It makes our values more explicit and constitutes a real-world effort to realize them through training. In addition, unlike evolution, the process of education can be quite intentional. We teach kids to share *because* we want them to be more cooperative. Viewed in this light, concerns about the absence of a link between the causal origin of our moral beliefs and their normative validity appear somewhat less compelling. When we consider the

---

[12] Mazzucato, 2017: 13-14.

formation of moral beliefs concretely, in terms of culture and education, the more general arguments for moral skepticism lose some of their sting.

This observation, reflected in some of the literature on evolutionary debunking of morality,[13] led me to consider the role of sociality in the formation of moral sentiment, and it fit nicely with my long-standing interest in the origins and legacy of German idealism. The authors that I will be dealing with in chapter 3-5 all had something to say about moral education. And, while obviously writing before Darwin, we will see that concerns about anachronism are likely to be overblown. Jean-Jacque Rousseau also believed that human beings have innate psychological capacities that help them navigate their environment. He also believed that these natural instincts can be harmful in the context of modern technology and large scale social environments. The difference, obviously, is that Rousseau thought these instincts were the perfect gift of God, not the refined yet imperfect consequence of a long history of evolution.

In chapter one, I will introduce the concept of evaluative perception and argue that it is of great importance to moral philosophy. I will argue that moral philosophers will need to appeal to evaluative perception if they want to defend what I call 'common sense moral objectivism.' It will be seen that common sense objectivism is compatible with some forms of realism and constructivism but not with the error theory. Issues of the language of moral philosophy are set aside. In chapter two, I detail three varieties of Darwinian debunking, arguing that none are sufficient to refute common sense objectivism but that selective evolutionary debunking arguments can contribute greatly to moral psychology and epistemology.

---

[13] Fitzpatrick, 2004, 2015.

In chapters 3-5, I will consider the views of Rousseau, Kant, and Hegel regarding the moral conscience in order to develop a view of the moral conscience that is compatible with what we know about how evolution shaped our moral sentiments. I highlight the legacy of perfectionism and the emphasis that these authors place on human freedom. These elements are already present in Rousseau's naturalistic sentimentalism, but they get developed within the German idealist tradition. Kant's superior account of perception makes his account of conscience more sophisticated than Rousseau's and capable of drawing a principled distinction between amoral passions and genuine moral sentiment. The result is a picture of the conscience that emphasizes humility and reverence for human freedom. Hegel's account of ethical life shows how social context, and its corresponding structures of meaning, generate moral content, while remaining sensitive to cultural difference and consistent with common sense objectivism. The resulting moral worldview is one that emphasizes freedom and respect in a socially embedded context.

1.   ON THE IMPORTANCE OF EVALUATIVE PERCEPTION

1.1.   Introduction

We rely heavily on the way things seem when making and justifying evaluative judgments. Intuition pumping thought experiments are common in normative and applied ethics, and the history of philosophy is loaded with vivid images intended to inspire and sharpen our moral responses. Appeals to the way things seem play an indispensable role in how philosophers talk about value. If Plato was a rationalist, it didn't stop him from treating moral inquiry like a beauty pageant. His description of the tyrannical man seems intended to inspire a response, as if the way it appears to the reader were an argument unto itself. We're just supposed to see that the tyrannical man's soul is corrupt.[14] We might find Kant's appeal to a non-partial rational *observer* surprising, given his refusal to derive duty "solely from experience." But this appeal is essential to his rejection of hedonism (**G**, 4:393, 4:407) Happiness is not good in itself, and we know this because we can be sure that a fair observer would derive no satisfaction from observing the happiness of a wicked person. What makes this argument convincing is our ability to approximate the conditions of such perception and concur with the verdict. Maybe it helps to have read stories or to have had an experience like this. But, in any case, we put ourselves in the shoes of the hypothetical observer by an act of imagination. Philosophers appeal to how things seem at different levels of their moral theory. The examples discussed so far are from metaethics, but one may prefer to make such appeals at the level of applied or normative ethics.[15] I'm not

---

[14] Plato, 1992: 361d, 540c.

[15] Peter Singer's (1972) thought experiment is a characteristic example from applied ethics. Intuitionists like Ross (1930) and Robert Audi (2005, 2015) treat the intuition of self-evident moral principles—like, say, that you ought to keep your promises—as the epistemic ground of normative ethics. Audi's distinction between intuition and perception is discussed below.

taking a position on the right *way* to appeal to observation. In this chapter, I only mean to argue *that* such appeals are crucial to theories of value.

In section 1.2, I will spell out what I mean by 'evaluative perception.' Then I will lay out some minimum conditions that I think any value theory should meet—a view that I call, 'common sense moral objectivism.' Common sense moral objectivism is compatible with moral realism and some forms of moral constructivism. With these limitations in place, I will argue for the importance of evaluative perception for realists about value who want to overcome the regress argument (section 1.5). Then I will argue for the importance of evaluative perception for constructivists who want to evade the empty formalism objection (section 1.6).

1.2. The Perception of Value

Evaluative perception is a general term for the way in which one can be *directly aware* of whether some thing, action, or state of affairs is good, bad, better, or worse.

Sometimes evaluative perception operates according to certain background assumptions, like when one notices that a tool is instrumentally fit or unfit for a given task. But, of particular interest are those cases in which something is taken to be good without qualification and without reference to anything else—like when a non-expert directly apprehends the value of a complex aesthetic experience.[16] If it is possible to be directly aware of what philosophers have called *the*

---

[16] Even experts tend to speak fondly of their novice experiences. In the introduction to her book on professional sommeliers, Bianca Bosker (2017) reports that most of the sommeliers she met had "a story about *the* bottle that launched their obsession with wine. Usually, Saul-on-the-road-to-Damascus moment via, say, a 1961 Giacomo Conterno Barolo sipped in a little restaurant in Piedmont, Italy, overlooking the Langhe hills, the beech trees swaying as a gentle fog curls up from the valley floor. It's something of a formula: Europe + natural splendor + rare wine = moment of enlightenment."

good—of what's *intrinsically* or *non-instrumentally* good—then these cases will surely be of special interest.

In this section, I will say a bit about why have chosen to use the phrase evaluative *perception*. Put quickly, I think that this perception-talk is especially illuminating, while also general enough not to rule out competing views. The question of whether and how we may be able to know the good cannot be decided by how we choose to define our terms, and since there may be multiple ways of drawing the relevant conceptual boundaries, I want to remain as general as possible in making the argument that something like evaluative perception is important for ethical theory.

1.2.1. Direct Deliverances and Cognitive Penetration

In an essay about the importance of intuitions in ethics, Robert Audi compares the faculty of intuition to that of perception:

> …just as we have perceptual faculties such as vision, we have intellectual faculties, including intuition—reason in one of sense of that wide-ranging term. The faculty of intuition is a kind of apprehension capacity: a non-inferential capacity through which we know what we intuitively do know. Intuition is like perception in the directness of its cognitive deliverances, but it does not require unresponsiveness to premises. It is not a modular part of the mind but a general capacity needed for philosophical reflection… It is appropriately called non-inferential because of the way it responds to content in context: above all, to the appearance of truth in propositional objects considered in their own terms rather than as supported by premises. *Intuition is like perception in the directness of its cognitive deliverances, but it does not require unresponsiveness to premises.* Intuitions (as deliverances of this capacity) may not come fast, any more than an aesthetic response to a complex painting need come at first look, but intuitions are not *inferentially dependent on premises*.[17]

---

[17] Audi, 2015: 60-61. Italics mine.

I might have called this part of the dissertation, "on the importance of moral intuition" were it not for fear of deciding too much too early on. Elsewhere, Audi describes his view as "Kantian intuitionism," and in the later chapters I will interpret the Kantian moral conscience in a way that (I think) is largely consistent with Audi's conception of intuition. In this dissertation, I am most interested that capacity for direct evaluative deliverances that have historically passed under the vague name, "conscience." Whether and in what way this capacity is responsive to premises can be answered in a variety of different ways, depending to a certain extent on other elements of a given philosophical system. Ontological questions may weigh heavily on the plausibility of given notion of conscience. Audi, for example, is a moral realist. So, it is likely (and, in fact, so) that conscience will function differently in his system than that of a constructivist. In Audi's work on moral perception, he makes it very clear how perceptual features—and even emotion--can serve as an accurate source of moral judgment.[18] But intuition has pride of place in his system as the faculty providing us with access to self-evident and properly basic moral principles. I'm making no argument against his system; I am simply allowing that there are other ways to carve the turkey. My thesis in this chapter, then, is that, however you slice it, conscience will need to play a role. I call conscience 'evaluative perception' mainly to stress its directness. It should be obvious that this doesn't rule out systems—such as Audi's—that use phrases like moral 'perception' for a more specific purpose.[19]

While we have seen that Audi distinguishes intuition from perception on the grounds that intuition needn't be unresponsive to premises, Susanna Siegel defends a view of perception that

---

[18] Audi, 2013: Chapters 5 and 6.

[19] Audi (2013) reserves the term perception for moral knowledge by experience, preserving his distinction between intuition and perception. As taxonomy, this is a good distinction to make, but Audi's Kantian approach complicates a strict division between empiricist and rationalist approaches in epistemology.

emphasizes its susceptibility to cognitive penetration. In *The Rationality of Perception,* she

draws our attention to what she calls *the problem of hijacked experience.* A "hijacked

experience" is one in which our expectations—often the word 'prejudice' seems especially

appropriate—can render judgments based on these experiences unreliable. Siegel thinks this goes

all the way to perception. To illustrate the point, she sometimes uses the admittedly apocryphal

example of 17th Century Dutch preformationists, who shortly after the invention of the

microscope, supposedly saw embryos in sperm cells.[20] This would be an extreme case of

hijacked experience, but Siegel's everyday examples seem quite plausible: a vain actor who

(mis)interprets the crowds boredom as approval; a woman who's afraid her friend is angry at her

and (mis)interprets his neutral expression as anger; and a white seventeen year-old boy from the

U.S. who harbors certain prejudices due to his lack of interaction with black people.[21] More than

these hypotheticals, Siegel is especially concerned with real world cases in which racial bias has

been demonstrated in snap judgments concerning whether one is carrying a gun or an innocuous

object.[22] Siegel's explanation for hijacked experience is the rationality of perception thesis,

according to which "perceptual experiences and the processes by which they arise can be rational

or irrational."[23] One way of understanding this is to allow that how we experience things depends

on features of our worldview that may be embodied in our attitudes and determines what we

notice. This world-view is the basis of inferences that we carry out without noticing. Thus, she

maintains that under certain circumstances we may be rational in trusting the deliverances of

---

[20] Siegel, 2017: xiv and 6-7.

[21] Siegel, 2017: 3-13 and 181-85.

[22] Siegel, 2017: 174-77. She cites, among others, Payne 2001; and Correll et al. 2002 and 2015; and James et. al. 2013 on weapon categorization.

[23] Siegel, 2017: 15.

experience. Perhaps it's possible to cash out hijacked experience purely in terms attention or background beliefs. I'm not deciding that here. But, to me, two things seem clear: (1) judgments based on perception can be deeply corrupted by facts about the perceiver; and (2) it is an open question—at least for my purposes here—whether and to what extent perception is inferential.

Opposite the problem of hijacked experience is the issue of expert intuition. This issue has been the focus of much of the literature on moral perception. For example, Watkins and Jolley call moral perception an "intellectualized perceptual ability," and they refer, by way of analogy, to "the mechanic who can tell what is wrong with the car by listening to it run." The mechanic, they say, "can hear something that the non-mechanic does not hear."[24] One might initially be tempted to believe that evaluative perception is necessarily something vague and mysterious, while everyday visual perception, say, is something concrete and easy to pin down. Expert intuition, shows that even mundane experiences are not as simple as they seem. Following Herbert Simon, psychologists tend to describe intuition as mere "recognition" or "knowing without knowing how one knows."[25] Examples, like Watkins and Jolley's mechanic, illustrate how perceptual abilities contain an intellectual component.

1.2.2. External Cues

This intellectualized conception of perception draws some empirical support from psychology. Gary Klein's work on expert intuition is a source of fascinating examples, and in a collaboration with Daniel Kahneman, the two even attempt to define the parameters under which expert intuition can be considered a reliable source of knowledge. In an earlier book, Klein

---

[24] Watkins and Jolley, 2002: 77.

[25] Simon 1992. Kahneman, 2011: 235. See also, Epstein, 2010: 296.

recounts an interview with a firefighter who told him about an experience in a burning building. After ordinary attempts to put out the fire by dousing failed, the lieutenant was struck by a feeling that something was wrong. He couldn't say, even at the time of the interview, what it was that he noticed, but something was off. Instinctively, he ordered the crew to leave the building. Moments later, the floor they'd been standing on collapsed. His gut response *one-knows-not-what* saved the lives of the members of his crew. In the interview with Klein, the lieutenant—a commander by this time—described the feeling as E.S.P.[26] In a collaboration with Kahneman, Klein attempts to identify the conditions under which this kind of intuitive judgment can be considered trustworthy.[27] They conclude that it is possible for individuals to develop the basis for reliable expert intuition, provided that the expert has the opportunity to learn the regularities of the environment in which their intuition operates. For example, chess masters develop the ability to recognize possible combinations and regular patterns in what Kahneman and Klein call a high validity environment. The rules are fixed. Although there are more possible combinations than anyone could hold in thought at once, chess masters develop a sense of what works based on unconscious recollection of what's has or has not worked in the past. Once one has truly mastered an environment like this, one "will not encounter challenges that are genuinely new."[28] This makes expert intuition largely reliable. Low validity environments, on the other hand, admit of frequent and irregular changes that undermine the reliability of such unconscious pattern recognition.

---

[26] Klein, 1999: 32.

[27] Kahneman and Klein, 2009. See also Kahneman, 2014: Ch.25.

[28] Kahneman and Klein, 2009: 522.

The role of the environment in assessing the reliability of expert intuition related to Siegel's discussion of cognitive penetration. We have seen that there is room for disagreement concerning the extent to which perception is *intellectual*. Perhaps there is a kind of spectrum with a kind of pure *raw sensation* view of perception on one end and an *intellectually loaded* view of perception on the other. But there is also room for disagreement concerning *where the action is*—so to speak—regarding perception. That is, were it not for Kahneman and Klein's important point about high and low validity *environments*, we might think that the most important part of expert intuition is what's going on inside the expert's head. But if the environment determines what patterns get internalized, and the continuity of the environment determines whether past patterns will repeat again, we may what it is exactly what role the subject and his or her unconscious is playing in this whole operation. A simple answer might be that the expert is creating a map of the world and acting according to that. So, when the map is accurate, the spontaneous intuition will fit the environment and the expert will act appropriately, even without consciously thinking about what to do (E.S.P.). But, at least sometimes, this probably isn't what's going on. Research on the "outfielder problem" suggests that when a person runs to catch a fly ball, that person is not simply predicting where the ball will land based on some internalized model of the world.[29] Instead, the outfielder adjusts his or her location relative to the ball by making sure that one's gaze, when continuously looking at the ball, continues to slowly rise at the same rate. This "continuous coupling of visual information and movement" doesn't require the outfielder to store and then access information about the ball's trajectory, only to keep his or her eye on the ball repeating behavior relative to it's current

---

[29] Fink et. al., 2009. See also Sloman and Fernbach, 2017: 96-101.

movement that has been successful in the past. Sloman and Fernbach use these sorts of examples

to help build their case that the brain ought not to be thought of as a computer processing stored

information. Rather, they maintain that, whatever computation may be going on, it involves

elements of the external world as well. On their somewhat radical view, thinking is *never* entirely

solitary and "the world is your computer."[30]

1.2.3. Intuition or Perception?

While it may seem that I am using these insights from Siegel and others to dispute Audi's

notion of intuition in favor of intellectualized moral perception, I want to insist on leaving this

sort of question open. There are multiple coherent ways of explicating the faculty by which we

have direct access to value. I will develop three in the second part of this dissertation. My

purpose in this chapter is to argue that *some* such account needed. What should be clear at this

point is that one way in which such accounts may differ is the extent to which perception is said

to contain and/or depend upon cognitive elements. There are also different accounts of value. A

workable value theory will need a conception of value that coheres with the account of

perception in a way that allows for moral knowledge. We will see that this is a problem for

Rousseau's sentimentalism. And the problem of making *the law of the heart* intelligible is a

central problem to which his theory of education is the solution. To adopt my terminology,

Audi's intuitionism is a theory of evaluative perception well-tailored to other aspects of his

overall ethical theory. Sarah McGrath rejects Audi's view by arguing that, "it is not clear that the

basic Kantian thought—that one could know moral facts by the exercise of one's general rational

---

[30] Sloman and Fernbach, 2017: 96. They acknowledge their debt to John Haugeland and his critique of "Good Old-Fashioned Artificial Intelligence," and it seems clear to me that their work is of a piece with the work of Richard Rorty and others who reject the representationalism and the "mirror of nature."

capacity—is correct." But a great deal depends on this phrase "general rational capacity." She

explains:

> being good at reasoning and in possession of the requisite concepts seems
> insufficient for being in a position to obtain moral knowledge, in the way
> that, arguably, being good at reasoning and in position of the requisite
> concepts is sufficient for being in a position to gain knowledge of logical
> or mathematical truths.[31]

From this it seems clear that McGrath is talking about what Kant would have called the faculty

of *understanding*, which is importantly distinct from the broader faculty of *reason,* and more to

the point, this characterization of reason is narrower than Audi indicates in a later article with his

phrase, "reason in the broadest sense of that term."[32] In part two, we will see that, properly

understood, Kant's view is grounded on a direct awareness of freedom and obligation; reason

furnishes its own moral content, which needn't be understood in the merely formal terms that we

typically associate with mathematics and logic. This view may be wrong, but it is more robust

than the charge of empty formalism suggests. McGrath isn't really criticizing Audi's view, only

the idea that reason—in the sense reserved for *a priori* truths of mathematics and logical analysis

—can be sufficient for moral knowledge. Both seem to hold that there is a kind of direct

awareness of (at least some) moral truths that is not derived from any non-moral evidence.

Jesse Prinz puts this point very clearly when he compares his sentimentalism to

intuitionism. On Prinz's view, what makes intuitionism so compelling is actually best explained

in sentimentalist terms:

> Intuitionists believe that moral judgments are self-justifying; they do not stand in
> need of independent argumentative support. In this respect moral judgments are

---

[31] McGrath, 2004: 223.

[32] Audi, 2015: 57.

like certain perceptual judgments or mathematical judgments. It is difficult to come up with arguments for the self-justification thesis, and indeed some intuitionists simply assert that it is obviously true. They seem to base this assertion on the phenomenology of moral judgments: moral judgments *seem* self-evident. I think sentimentalism can explain this phenomenology. Sentimental judgments generally seem self-evident… emotionally grounded moral judgments have a kind of perception-like immediacy that does not seem to require further support. We can feel that killing is wrong. Indeed, far from opposing intuitionism, sentimentalism offers one of the most promising lines of defense.[33]

So, to the extent that appeals to *intuition* and appeals to *sentiment* and appeals to *emotion* and appeals to *self-evidence* all seem to involve an appeal to some kind of *direct apprehension* of a thing or state of affairs being good, bad, better, or worse, I feel justified in calling it by the name 'evaluative perception'—if only to emphasize what Prinz calls its "perception-like immediacy" along with (we will see) its crucial role in ethical and evaluative theorizing.

1.2.4.   Summary

Now, we have seen that the varieties of evaluative perception admit of some important distinctions. There is the question of *cognitive penetration*. Some versions of intuitionism are so thoroughly rational that we may hesitate to use the word 'perception' (Audi's intuitionism). But as long as these involve the direct apprehension of value, they meet my definition of 'evaluative perception.' Other accounts are more directly comparable to physical sensation (Prinz's sentimentalism). In between, we have intellectualized accounts of perception (Siegel's rationality of perception thesis). There is also the issue of *experts* and *novices*. The question of the extent to which training and education makes us better (or worse) at moral perception will be central to this project. There is also a distinction between the perception of *basic* and *derived* moral truths. Audi maintains that we have access to self-evident moral propositions by intuition, while moral

---

[33] Prinz, 2006: 37.

disagreement consists in both disagreement *on reasons* and *in reasons*.[34] That is, philosophers

may disagree *on* the moral principles that justify their moral judgments, but people also might

disagree *in* their moral judgment regarding, say, a particular instance of wrongdoing. *Derived*

moral truths depend for their validity on more *basic* moral truths. For example, someone might

immediately recognize that there's something blameworthy in arbitrarily kicking a pregnant dog,

But that same person might not know precisely what it is that makes it so. Maybe this person

recognizes that cruelty is bad *and* that some kinds of living beings have rights. Then, despite

simply being aware that it's wrong to kick the dog, this person may not know whether it's

because: dogs have rights; that kicking dogs is cruel; some combination of the two; or some third

thing. So, evaluative perception might be the direct apprehension of fundamental truths about

value, but it also might be the direct apprehension of value in particular instances. Evaluative

perception can play many different roles in the various kinds of moral theories, but it's hard to

see how a complete moral worldview could do with out it.

1.3.   Common Sense Moral Objectivism

Studying anthropology is a good cure for cultural arrogance. Customs and practices that we take

for granted are not as universal as we tend to think they are. Open-mindedness about different

ways of life is usually a good attitude to have, and we should be suspicious of those who point to

their own habits and customs and assert that everyone else ought to do the same. Jonathan Haidt

speaks approvingly of the everyday sacredness that he experienced during his time doing

---

[34] Audi, 2015: 76.

research in India.[35] By studying other cultures, we can learn about ourselves and about how

others throughout the world find meaning in their lives.[36] It is a mistake to take radical diversity

as proof of cultural relativism. While it's one thing to take off or your shoes when entering the

house, some of our evaluative commitments concern profound injustices and abuses—

cannibalism is a classic example.[37] Still, accepting this distinction, one may remain tempted by

the easy relativism of: *who am I to say?* If other people live in radically different ways from my

own, it remains to be decided what authority I have to say that one of these ways is better than

the other. Struggling to find a good answer, I might (mistakenly) conclude that the answer is 'no

one.' This is further complicated by the fact that anthropology is usually thought to be an

empirical science.[38] As such, it may seem off-limits to appeal anything outside the ordinary

causal order of things, like moral facts or values. Evaluative judgments about the superiority or

inferiority of customs may seem supernatural or just out of bounds from an anthropological

perspective. It becomes tempting to say that there are two distinct but equally legitimate

standards by which people judge their own practices. To avoid lapsing back into cultural

arrogance, one is tempted to endorse cultural relativism. In this section, I will argue that this is a

mistake for three reasons. (1) The choice between ethnocentrism and relativism is a false

dichotomy. (2) At least some values are not culturally relative--in particular, those concerning

---

[35] Haidt, 1996: 85. "I adopted the Indian practice of removing my shoes at the door, and asking visitors to do the same, which made my apartment feel more like as sanctuary."

[36] Tim Ingold (2013: 3-6) takes an extreme view. He argues that anthropology *just is* this sort of engaged cultural understanding. Anthropologists, he thinks, should limit themselves to a kind of engaged participation within a culture. Through such participatory learning, the anthropologist allows him or herself to be transformed. Data collection and description belong to a separated discipline, ethnography— which he seems to also find morally questionable. While there are certainly profound insights to be gained from participant observation, I find his case for restricting anthropology in this way unconvincing.

[37] Montaigne, 1953: 1.30. Prinz, 2012, 322-24.

[38] See note 23 for an alternate view.

respect and toleration for cultural difference. (3) In addition, many writers working on the empirical investigation of human nature (anthropology, psychology, etc.) implicitly endorse a view that I will call common sense moral objectivism.

Let's take these reasons in reverse order, starting with some ethical observations from those who investigate human nature empirically. I include representatives from anthropology, evolutionary psychology, and the social sciences, but my point is not to state what *all* empirical investigators of human nature believe. There is surely a broad divergence of views. No doubt, some psychologists endorse error theory; no doubt, some give little thought at all to questions in ethics and philosophy. That's fine. My point is simply to identify what I take to be some representative passages that indicate some widespread moral points of view. Some social scientists are quite explicitly engaged in ethical debates. Proving that scientists believe in moral truths has gotten easier lately, thanks to authors like Steven Pinker and Sam Harris who go so far as to defend their moral views publicly.[39] In their popular writings, evolutionary biologists take pains to disassociate from the history of genocide and social Darwinism. One approach is to argue that scientific facts can't determine values.[40] Another is to appeal directly to moral principles. Now, the mere fact that some smart folks in the natural sciences believe in some objective moral principles is not proof that any such principles are true. In the next chapter, we will see that some philosophers think that our ability to know moral facts is seriously undermined by evolutionary accounts of human nature: the skeptical argument becomes

---

[39] Pinker, 2018. Harris, 2012.

[40] Henrich, 2016: 96.

especially interesting when the same scientists who provide the premises have reasons to fear the conclusion.

The study of other cultures can inspire cross-cultural understanding. Indeed, the values associated with such understanding are reflected in the works of many prominent social scientists. In their book on cultural evolution, Robert Boyd and Peter Richerson say the following:

> Peoples and their cultures are wondrous and diverse. The study of human diversity highlights how much humanity we share with the most exotic of our fellows…We don't subscribe to an extreme form of cultural relativism (Nazism, after all, was not quaint German folklore). However, the anthropologists' practices of refusing the easy pleasures of ethnocentrism in favor of reserving judgment about others societies—at least until you understand them well—has much to recommend it.[41]

This passage illustrates two features of what I'm calling common sense objectivism. First, there is the respect for cultural difference. The authors are not describing best practices for anthropologists working in the field—although, being respectful may be useful in that context too. They are saying that there is something morally wrong about imposing your own cultural practices on others simply because you prefer your own. Second, the parenthetical reference to Nazism and German folklore suggests that there is a distinction to be drawn between cultural differences that should be respected and those that constitute monstrous ethical violations. This distinction may be explained in terms of human rights, but it doesn't have to be. There is plenty of room for debate concerning the nature and scope of moral wrongs, but what's clear from this passage is that the authors acknowledge a distinction between morality and mere custom. Some actions are simply not acceptable, regardless of what anyone's culture says about them.

---

[41] Boyd and Richardson, 2006: 255.

This idea is also reflected in notions of moral progress. Boyd and Richerson approvingly cite Charles Darwin's remarks on the treatment of slaves in Brazil. They quote a passage from his journals condemning slavery and praising Britain for freeing the slaves in all its colonies in 1838:

> It makes one's blood boil, yet heart tremble, to think that Englishmen and our American descendants with their boastful cry of liberty, have been and are so guilty: but it is a consolation to reflect that we have made a greater sacrifice than any nation to expiate our sin.[42]

This condemnation of slavery reflects the religious language of guilt, sin, and redemption through sacrifice, but even if we dismiss his moral language as rooted in Darwin's Western cultural context, I doubt many would disagree with his main point. Darwin observes that Britain is better for having abolished slavery. If cultural relativism were true, such an assessment would be nonsense. Suppose relativism were true. Then, as long as slavery was approved of prior to abolition and disapproved of post abolition, the practice of slavery prior to abolition is equally legitimate to the refusal to practice it afterwards. In both cases, the practice (or refusal to practice) is consistent with the norms set in place by the cultural context. If relativism is true, there is no independent system of evaluation to which both may be subjected. Darwin's reference to the "cry of liberty" makes little difference to my point. Surely, slavery was inconsistent with the values of liberty and justice. But if all systems of cultural values are equally legitimate, it makes little difference if a system holds values that are inconsistent—unless, that is, we are willing to impose upon them the value of consistency, which would involve appealing to something objective. To claim that progress (even progress on one's own terms) has been or can

---

[42] Quoted in Boyd and Richardson, 2006: 283n40.

be made requires the willingness to assert that some ways of doing things really are better than others.

   I call the moral worldview that I find reflected in these passages common sense moral objectivism. 'Objectivism' refers to the fact that *at least some* moral values are objective. That is, the view requires us to say that sometimes one set of cultural values *really is* better than alternatives.[43] Systems of slavery really are worse than more humane divisions of labor and goods. These kinds of evaluative judgments do not always come down to cultural preference. That a system of social organization is somehow broken, inadequate, or unfair is something that we are sometimes able to tell. We very often get it wrong, but it is enough that we sometimes get it right. It follows from this that value is not entirely culturally determined. At this point, it's enough that most people agree that our common sense judgments about things like slavery and genocide are correct. The philosopher's job is to go deeper and attempt to identify the subtler contours of apt moral judgment. 'Common sense' refers to the fact that (I think) very few people would disagree with what has been said so far in this section; it also refers to the fact that the view is consistent with a wide range of views about complex philosophical issues. Later in this chapter, I will argue that common sense objectivism is consistent with moral realism and most forms of constructivism. Among the latter I would include some forms of social contract theory, insofar as the idea of contractual obligation sets objective limits on how a society can be organized. In later chapters, we will see that the salient evaluative features of human action become more perspicuous when actions are imbedded in their unique cultural context.

---

[43] Good examples can be found in: Churchland, 2011: 196, 200; Pinker, 2011: 622-654.

Given the diversity of their topic, it's not surprising to find that those who study human nature have important things to say about tolerance and cross-cultural understanding. Often the lesson that comes from study of moral psychology is that we need to beware of bias and our tendency to moralize about others. This seems to be where those who study human nature scientifically are most comfortable making value judgments. Recent popular books by psychologists and cognitive scientists that venture into discussing moral and political issues, stress the need to overcome cognitive biases in politics, morality, and even business management. These authors tend to emphasize rationality and fairness, sometimes even explicitly invoke notions of moral objectivity.[44] Regarding the latter, Steven Pinker puts the issue bluntly: "the world has far too much morality."[45] Properly understood, this is a value judgment with which we should all agree. In saying that there is too much morality, Pinker is referring to the way in which we tend to moralize about things that are not appropriately moral. In discussing the psychology of how and why we moralize, he shows how it magnifies the risk of harm. We can become hostile toward those we find different and rationalize our hostility as if it were just punishment. The claim that there is too much morality might be restated: there is not enough understanding. The tolerance and openness that we find reflected in the attitude of those who engage in the scientific study of human nature has less to do with relativism than the Delphic dictum, "know thyself." But that we can improve society by promoting greater understanding implies that there are at least some objective standards concerning how society could be better.

---

[44] Recent examples include: Haidt, 2015; Bloom, 2016; Pinker 2018; and Sloman and Fernbach, 2017: 171-94.

[45] Pinker, 2011: 622.

The move from the scientific study of human nature to relativism oversimplifies the debate concerning the nature of value. The argument, to the extent that there is one, seems to go like this. If there were objective standards for the goodness of *x*, then all good *x's* would be the same. There is a diversity of good communities. Therefore, there are no objective standards for what makes a community good. The conditional is obviously false. If it were true, there would be no objective standards for what makes a good or bad game. Chess is a fine game. Sometimes, for an interesting variation, players will tweak the rules by moving the starting position of some pieces or tinkering with how opposing pieces may be captured. These are often fine games too. But imagine making it a rule that the tallest player could change the rules at any time arbitrarily. As the taller player's king is one move from checkmate, he or she changes the rules so that the winner is the player who has the fewest pawns. This would be an awful game. The existence of a plurality of ways in which something can be good does not rule out the existence of objective conditions of that thing's goodness. There may be a plurality of potentially conflicting values that any good system of social organization must reasonably accommodate. Or, perhaps, there is some single absolute value that may be accommodated in a plurality of ways. In either case, the diversity of good societies does not rule out the kind of objectivism that I am advocating here. Common sense objectivism is perfectly compatible with what we might call 'pluralism.' And pluralism has the advantage over relativism that it can respect diversity without denying that some systems of social organization are objectively better than some others.

1.4. Moral Ontology and Critical Force

It may be possible to develop a moral theory that does justice to common sense moral objectivism without making any appeal to the broad notion of evaluative perception discussed in section one, but I am skeptical about the prospects for this. At the heart of the problem is question of how we have moral knowledge. And how we have moral knowledge depends in part on *what moral truth is*. It seems to follow from common sense moral objectivism that true moral judgments will have some degree of *critical force*. That is, they will hold regardless of the tyrant's desire for them not to. They tyrant may still have the power to do what's wrong anyway but not the power to change the truth about its value. Slavery was wrong, even when slaveholders had the power to do it anyway. To do justice to common sense objectivism, therefore, what's needed is a moral *ontology* that does justice to critical force and an epistemology that accounts for how we have access to whatever moral truth is. In this section I will sketch the standard taxonomy of moral ontologies, ruling out what philosophers call the *error theory* and explaining my reasons for setting *expressivist* views to the side. Then, in the remaining sections, I will argue that moral realism and constructivism, in order to be viable, need to give a solid account of evaluative perception.

To get started, suppose that Allen observes Betsy doing a good deed for Cedrick. If Allen says, "Betsy did a good thing" we can interpret the predicate 'good' in a variety of ways depending on how we understand the meaning of moral language and the ontology of the property *goodness*. In this dissertation, I will largely ignore the issue of moral language. Some authors argue that moral claims are not intended as descriptions at all. Instead, the claim 'Betsy

did a good thing' is better interpreted as an expression of approval (*Expressivism*).[46] 'Murder is wrong' is interpreted as an expression of disapproval. Whether this is the correct interpretation of moral language is not the issue I am dealing with here; whether an action is wrong depends on whether it meets the conditions of an action's being wrong. If murder meets those conditions (whatever they are), I will take 'murder is wrong' to be true. If not, I'll take it to be false. Even if these authors are right about how we typically speak, I see no obstacle to posing the question as if it were a description and, hence, subject to standards of truth or falsity. It may be more comfortable for some to say that Allen's expression is 'apt,' rather than 'true.' That's fine. For our purposes, what matters is whether it is knowable and whether it holds independently of mere preference.

Among those who take moral language to be descriptive, there is the division between realists and anti-realists. *Realists* maintain that evaluative judgments are true descriptions of mind-independent facts that are agent-independent.[47] Anti-realists can be subdivided into two main categories on the basis of these two conditions. There are those who deny that evaluative judgments ever refer to *true* descriptions, maintaining instead that all moral claims are false. On this view, moral language is literal, but there are no moral facts. Following Mackie, we'll call

---

[46] The classic version of this view can famously be found in Ayer, 1952. For a more sophisticated modern approach to the view, see Blackburn, 1998.

[47] My taxonomy here is not intended to depart in any significant way from the standard view: See Sayre-McCord, 1988. Later on, however, I will raise some doubts about the significance of the distinction between realists and constructivists.

this view the *Error Theory*.[48] Then there are those who allow that some evaluative judgments are true descriptions, but deny that the things they are describing are agent-independent. On this view, moral facts are facts about agents. The latter view is called *Constructivism*. I will deal with realism in the next section and with constructivism in the following section.

The arguments that I will address in the following chapter are typically associated with the *Error Theory*. So I will ignore debates over moral language and use the language that error theorists use. In order to treat anti-realists and realists as interlocutors, it's necessary to suppose that they're talking about the same thing. Beginning with the assumption they may be talking about different things altogether makes it difficult to bring them into conversation. I will suppose that terms like 'good' are intended objective descriptions and ask about whether they are true and how we know them.

One of the main puzzles concerning moral perception is the question of how it can be possible to perceive something as odd and apparently immaterial as value.[49] The great advantage of the error theory is that there is no special mystery as to *how* Allen sees that what Betsy did was good. He doesn't! He sees something that he finds agreeable, and he wrongly infers from this that it is good. Why it is agreeable to him is a psychological question. Perhaps Allen belongs to a species with an evolved tendency to take pleasure in certain kinds of helping behavior. While we should be careful not to turn the error theorist into a straw man, the view is clearly

---

[48] Authors sometimes use the word 'nihilism' for the view that there are no moral facts, but I have tried to avoid this term for historical reasons. For example, Friedrich Nietzsche clearly denies the existence of moral facts (*Werke IV:* 979), and sometimes seems to use the term nihilism to refer to the view that "finds the nothing behind all human ideals" (*Werke IV:* 1007). But for Nietzsche, this nihilism belongs to a more holistic picture of modern decadence that can't be completely separated from related features (symptoms) of this way of life: pessimism, cowardice, exhaustion (e.g., *Werke IV:* 947, 1008-17). F.H. Jacobi uses the term 'nihilism' to describe a combination of atheism and determinism that he associates with Spinoza (On Jacobi's relationship German idealism, see: Beiser, 1987 and DiGiovanni, 1992).

[49] Mackie, 1977. Harman, 1988.

incompatible with common sense moral objectivism. Error theorists needn't deny that Allen has

reasons for approving of Betsy's behavior. That Betsy likes to help out is reason enough for

helping Cedrick, and that Allen likes to see folks treat each other nicely is reason enough to

approve of niceness. But these reasons don't carry any objective weight. They give the agents

themselves reasons for acting, but they provide no means for adjudicating between agents with

differing preferences. The assertion that a society that prohibits slavery is better than a society

that does not turns out to be false, and any reasons a society has for approving of a slavery-ban

only hold for those who share a set of relevant preferences. Error theory, therefore, lacks the

critical force demanded by common sense evaluative objectivism. A moral theory must be able to

justify the common-sense intuition that extreme injustices, like slavery and genocide, are wrong.

If it can't do this, it is a sign that the theory has gone seriously wrong.

Constructivists and realists agree that some moral propositions are objectively true. In

asserting agent-independence, realists make it clear that true moral claims hold regardless of

anyone's mere preference. If the realist is correct, there is no reason to worry that 'genocide is

wrong' holds only for societies that share the view that genocide is wrong. Those who fail to

condemn genocide are straightforwardly wrong. Constructivists may say the same thing, but the

fact of the matter is agent-dependent. This raise two important questions: (1) is the wrongness of

extreme injustices like genocide, so to speak, *up to* the agent? And if not, (2) what exactly does

agent-*dependence* amount to? If we answer question one in the affirmative, it seems that the

constructivist has failed to establish critical force—it's hard to see how the constructivist could

even call this objectivity. If we answer 'no' to question one, then we need to clarify the notion of

agent-*dependence*. 'Dependence' is an ambiguous term. In a trivial sense, most moral facts are

'dependent' on moral agents for their truth-value because they are *about* moral agents. This is the same sense in which the truth of facts about the external world depends on the external world being a certain way. Under this interpretation of the agent-dependence of moral facts, even natural law theory would be constructivist to the extent that it derives normative principles from facts about human nature.

In another sense, moral facts 'depend' on agents for their truth-value because they derive their truth *from* the agent(s). On some versions of this view, the truth of a moral principle depends on the agent because it is logically derivable from an agent's more basic moral commitments.[50] This sort of view faces two problems. First, if moral truths are derived from the agent, it's not clear why they should be bound to what logically follows from their deepest moral commitments. Is the moral fact that one ought to do what logically follows from their more basic commitments agent-dependent in the same way? Second, an agent who is willing to revise his or her moral commitments all the way down could justify injustice.[51] So, it seems again that the wrongness of injustices like slavery and genocide is *up to* the agent. Another version of this view, strikes me as more promising. This attempts to derive moral principles from the conditions of rational agency itself.[52] If successful, this strategy grounds moral principles in a way that is valid for all rational agents. This approach does not leave moral principles up to the agent, but it does seem to rest on the same principle of consistency as the other constructivist view. In any case, my concern here is with critical force. And it seems likely that both realist views and some

---

[50] For example, Street: 2016.

[51] Chapter Five considers this sort of approach in more detail.

[52] Korsgaard, 1996a. See chapter five for a discussion of the self-conception strategy.

constructivist views can establish the wrongness of sever injustice in a way that will satisfy the common sense moral objectivist.

1.5.   Realism and the Regress Argument

Realists maintain that at least some evaluative judgments are objectively true and that their truth is agent-independent. So, for example, if Betsy did something good for Cedrick—say, by helping him load the groceries into the car—then Betsy's action is good regardless of what anyone believes about it. To say that something is good is to make an objective judgment about a mind-independent fact. To perceive that something is good is to be perceptually aware of an objective and mind-independent value. Now, this perceptual awareness may not be the fundamental ground of the things goodness. When Allen sees that what Betsy did was good, it needn't be the case that Allen understands what makes this particular example of helping behavior good. The correct account could be something like this: kindness is intrinsically good and helping Cedrick with the groceries is an instance of kindness. Since helping in this way is kind, Betsy's action is good. But if, the two lived in a society in which grocery helping was taken to be condescending and disrespectful, the act in question might cease to be kind and, therefore, cease to be good. This possibility does not imply that the act is good *just because* the culture holds a certain view. The action is good because it's kind, and kindness is good regardless of the view the culture holds. What counts as kind can admit of cultural variation, while the value of kindness remains objective and agent-independent. On this description, kindness is treated as *basically good*, and judgment that the helping behavior is good is *derived* from the goodness of kindness and some facts about the given situation. Basic evaluative principles require no further inferential support.

Derived evaluative principles are inferences from basic evaluative principles combined with other facts, which may or may not be agent-independent.

The realist will need to maintain that there are *some* basic evaluative truths—at least one—that is agent-independent. That agent-independence is critical can be seen by considering the *desire theory* of value. According to this view, goodness just is whatever is desired by the agent. To identify "x is good" with "S desires x" or "S believes that x is good" would undermine realism's objective critical force. This isn't to say that evaluative truths can't be about agents. The claim, "all agents have intrinsic value" is agent-independent in the relevant sense, since its truth doesn't depend on the agent's preference.

So what role, if any, should evaluative perception play in the realist view? Evaluative perception could be the perception of basic evaluative facts. Maybe the goodness of fidelity or the badness of harm is self-evident. But evaluative perception could also refer to the immediate awareness of the goodness or badness of particular acts. For example, it may be determined upon reflection that Betsy's action was good because it was kind, even if Allen's perception of the goodness of the act involves no explicit reference to kindness. He might just recognize that the action was good, without being able to articulate why. Then, in cases of veridical evaluative perception (on a realist view) the phenomenal character of the experience is a sign that some part of what's being perceived can be expressed as an agent-independent evaluative truth that is either basic or derived from a basic evaluative truth. Often this phenomenal character is affective, a positive feeling about its object. The perceiver may see the non-evaluative features of the perception in an affectively toned way. A realist account of evaluative perception can take the affective character of the experience as evidence of the goodness of the object of perception. But

it needn't be this way in all cases. One may also infer that what is seen is good from the non-evaluative features of the perception and a basic evaluative principle. In this sort of case it may just be that the perceiver recognizes the action as an instance of an act already judged to have a certain value.

The realist doesn't just want to say that *there are* objective values. The realist also wants to be able to say that he or she knows what some of these are. The critical force of realism wouldn't be worth much if it bottomed out in the view that at least some moral claims are wrong but there's no telling which![53] The importance of evaluative perception is mainly epistemological. At some level, the realist will need to appeal to evaluative perception, or else there is little hope of overcoming what is known as *the regress argument*.[54] Since the reasoning behind this argument is very similar to the reasoning behind the argument above that realism about value presupposes some 'basic evaluative truths', I will use the term 'epistemically basic belief' to refer to a belief which is justified *without* reference to any other belief or set of beliefs. The term 'epistemically derived' will refer to beliefs that are justified without reference to any others. From here, the argument goes much like the one above. If any beliefs are to count as knowledge, then they are either epistemically basic or epistemically derived. If a belief is epistemically derived, then it depends on another belief for its justification. So, if there are any beliefs that count as knowledge, then there is either an infinite regress of justification or at least one belief is 'epistemically basic'.

---

[53] Joyce, 2006, treats this view as a version of error theory, which he refers to as moral agnosticism. "The metaethical label 'error theory' – coined by its proponent John Mackie – is usually reserved for the atheistic moral skeptic, but there seems nothing to prevent us from extending it to the skeptical position favored here: that no moral judgments are epistemically justified" (223).

[54] See Schafer-Landau, 2003: Chapter 19

There are three standard strategies for dealing with the regress argument: coherentism, self-evidence, and reliabilism. Coherentism denies the need for an epistemically basic belief and insists that a theory that hangs together requires no further justification. This strategy raises some additional questions. First, is it an epistemically basic belief that a theory that hangs together requires no further justification? If so, then two radically different—yet, internally coherent— systems might be asserted as basic. This could threaten realism's critical force. What could be said to the slaveholder who was willing to adjust other feature the moral web of belief in order preserve consistency with the permissibility of slavery? Perhaps some practices cannot be made coherent, but the immorality of slavery doesn't seem to conflict with mere logical consistency. But if slavery is logically incoherent, what shows that logical incoherence corresponds to moral wrongness? Evaluative perception may still be needed in linking the two—I will say more about this in the next section. Presumably, a more robust notion of coherence will be required, maybe one that can explain why greed is inconsistent despite appearing not to contradict the laws of logic. If we adopt a stronger notion of coherence, we would need to appeal to a more epistemically basic principle to determine which notion of coherence is the appropriate one to use. In which case, the principle of coherentism wouldn't be epistemically basic; it would be derived from this other principle. If it isn't epistemically basic that coherent views are justified, then we can ask what it is that makes one coherent theory justified and another not. Whatever new principle we came up with for comparing coherent systems, then, might be treated as epistemically basic. If so, we have abandoned coherentism by positing to some other epistemically basic belief. So, for example, perhaps what makes one coherent system better than another is that it can capture a greater diversity of truths with less dissonance. Then, if two

coherent systems are in conflict, perhaps a coherent system that can resolve the contradictions in

and between those other systems in such a way that preserves what's best in both at minimal cost

is inherently superior to the previous two systems. But to make this sort of a system work, a

particular kind of *experience* would be required to break out of the coherent system – the

experience of *recognizing* what's worth preserving in the conflicting systems. So, unless the

realist is willing to chalk up entirety of value-theory to the mere logical consistency—in which

case, giving up realism's critical strength—the realist will probably want to make an appeal to

evaluative perception in order to justify a preference for one logically consistent system over

another. In addition to this, an appeal direct apprehension would also be helpful in defending

coherentism from the skeptic. Why prefer coherence to disorder? There is an opening here to

make an appeal to self-evidence.

Appeals to self-evidence introduce propositions that are justified without reference to

anything else – i.e. epistemically basic propositions. But this raises the question of how one

knows which propositions deserve to be treated as epistemically basic. How does one recognize

that a belief is self-evidently true? The most straight-forward way to answer these questions is to

appeal to some type of intuition. For the realist about value, this will presumably take the form of

an intuition that some such thing or kind of thing is objectively good or bad. For example, if it is

self-evident that kindness is inherently good, one recognizes this by intuiting that kindness is

inherently good. This intuition doesn't necessarily have to be evaluative perception, but it seems

to me very likely that it will be. Perhaps the intuition of self-evident evaluative truths is similar

to the intuition of self-evident propositions in logic or mathematics, but it would be surprising if

objective values were intuited without the affective coloring that is the mark of evaluative

perception. While it is possible that there are self-evident objective values that can be known without appeal to any kind of evaluative perception, without supposing that we have a reliable capacity for evaluative perception, how these can be known remains a bit mysterious.

Reliabilism is the view that a belief is justified if it is formed in the right sort of way. There are many different candidates for reliable sources of knowledge about objective values, but a reliable capacity for evaluative perception is a strong competitor. If human nature is equipped with a reliable capacity for evaluative perception, the discovery of objective values could proceed in two ways. Perhaps, we begin with the most basic principles and work up to particular cases. On this picture, reliabilism looks similar to an appeal to self-evidence, but it has a crucial advantage. One perceives fundamental truths about objective value, and evaluates particular cases accordingly. But these truths themselves are not epistemically basic. They are fallible. What's epistemically basic is that beliefs formed by means of the capacity for evaluative perception are reliable. If conflicting evidence is discovered by means of this capacity, then even (supposedly) fundamental truths can be revised. The reliabilist can also go the other direction. The capacity for evaluative perception could be used to render verdicts in particular cases. Then one goes to work inductively, formulating general principles form reliable judgments about particular cases. Once general principles have been established, you might go back and look at tricky cases and attempt to resolve any contradictions. On this picture, reliabilism seems to have much in common with coherentism.

Positing a reliable capacity for evaluative perception can help the realist overcome the regress argument. If we grant that human beings have a conscience and that it's a good guide to telling right from wrong, then the realist can avoid the regress argument. While reliabilism is the

most direct way for a capacity for evaluative perception to overcome the regress argument, it seems likely that coherentist and self-evidentialist strategies will have a place for evaluative perception as well. And while one may wish to go the reliabilist route without positing a capacity for evaluative perception, I don't think these are likely to fare well. One contender would be an appeal to a reliable capacity for pure reason, but this suffers from the same problem as appeal to logical/mathematical self-evidence. There seems to be nothing evaluative in such intuitions. If the evaluative character of intuitions about objective value is going to break through to consciousness, the affective tonality that accompanies evaluative perceptions is required. So, if realism is going to overcome the regress argument, it seems very likely that it will need to posit some reliable capacity for evaluative perception.

1.6.   Constructivism and the Empty Formalism Objection

Constructivism is a form of anti-realism that attempts to preserve a notion of objective moral truth. According to the constructivist, normative standards are objective, but they are not agent-independent. I must admit that this is a view that I find somewhat confusing. I take objectivity to be the minimal requirement for critical force, but I'm not sure how much it matters that moral principles be agent-independent. Moral principles must be objective, in order for us to legitimately assert that extreme injustices like slavery and genocide are wrong. But what does it mean to say that such claims are both *objective* and *agent-dependent*? While the principle, all agents have value, is agent independent in the appropriate way, the principle, something is valuable just in case some particular agent values it, is not. The difference between these two principles is that the latter depends on what some agent (or group of agents) thinks or feels about

the principle. That it is true is agent-dependent. Whereas, the former principle is true regardless of what anyone thinks about it. According to constructivism, the truth of an evaluative principle depends on some subjective or inter-subjective feature of the agent. However, this agent-dependence is not thought to reduce value to mere preference. The constructivist believes it is possible for agent-dependent facts to be meaningfully objective and set serious limitations on what counts as good.

It's worth pausing for a moment to point out, while I have focused primarily on questions intrinsic value (*the good*), constructivists often limit their discussion to *the right*—or, what is morally permissible to do. It turns out that it is very hard for the constructivist to develop any kind of robust notion of goodness. Some Kantian constructivists, like Onora O'Niell, think that it can be done, but others, like John Rawls, are somewhat skeptical.[55] Because of these difficulties, many of the examples that I use when discussing constructivism will be more directly applicable to moral permissibility, even though I am ultimately more concerned with the prospects of a constructivist theory of value.

The constructivist strategy threads the needle between realism and strong anti-realism. The realist says that value is both objective and agent-independent. The strong anti-realist says that value is neither objective nor agent-independent. The constructivist says that value is objective but not agent-independent. So ,the trick for the constructivist is to develop a principle that is agent-dependent but does not bottom out in mere preference. That all agents have value is a realist principle. It is objective, and it is agent-independent in the relevant sense. That whatever the agent approves of is good is a nihilistic principle. It is not objective. The constructivist

---

[55] O'Neill, 1992 and 2002.

strategy is to attempt to place meaningful objective constraints on what agents can approve of without invoking an agent independent principle. This challenge becomes even harder if the constructivist wants his or her view to have the kind of critical force that makes realism an attractive view. The constructivist might try to derive objectivity instrumentally, but it's hard to see how this view has any advantage over strong anti-realism. The error theorist doesn't deny that there are more or less effective ways of satisfying one's preference; the nihilist denies that there is any objective measure for evaluating preference in themselves. If Jones says that slavery is good, the nihilist can try -- just as well as anyone -- to show that this is a bad fit with one or another of Jones's related commitments. But if Jones's views fit, or if he's happy to embrace views that don't quite fit, there is no claim of objective value to which the nihilist can appeal.

The constructivist needs to be able to say more, without saying too much more. Like the coherentist, the constructivist might impose a consistency criterion. It's not enough that Jones is happy to contradict himself. But then one might ask whether this is an agent-independent criterion. If so, the constructivist is a realist in disguise. If not, the constructivist should be able to say something about how such a criterion simultaneously *depends on* and *constrains* the agent.

In a game, there is usually no problem changing the rules a bit, provided that all the players agree. When a group of people decides to play a game the rules of the game ultimately *depend on* them, since they decide to play the game and could (presumably) decide to stop at any time. However, they are also *constrained* by the rules as participants in the game. On the model of a game, we can see how it is possible for an agent to be constrained by some criterion that also depends on that same agent. Of course, the agent needn't continue to play the game. Realists think that objective values set standards that hold come what may. There is no *choosing to*

participate in a system of objective values; there *just are* objective values. One kind of

constructivist says that there are some fundamental rules that the agent must endorse in order *to*

count as an agent at all. These fundamental rules place *universal* constraints on all agents,

despite also being agent dependent. If there are such constraints, then whatever they rule out

must be considered wrong for all agents. If they rule out slavery, then the constructivist will have

something to say to Jones. Something like, it is universally wrong for any agent to enslave

another and that a society that forbids slavery is objectively better than a society that permits it.

The constructivist can say this without positing the existence of any agent-independent values.

In order for constructivism to preserve the critical force that made realism so attractive,

the constraints placed on subjects must be universal. To be universal, these agent-dependent

constraints must follow from some fundamental feature of agency in general. For this reason,

constructivists often adopt the Kantian strategy of identifying *conditions of possibility*. If some

feature *F* can be shown to be the condition of the possibility of some thing or state of affairs *x*,

then for any *x*, *F* holds universally -- (*x*)*Fx*. If, for example, consistency is the condition for the

possibility of agency, then for any agent, that agent is consistent. This strategy was used to great

effect by Kant in the first critique. By showing that a principle of succession in time according

the law of cause and effect was required for our everyday perception of events, he was able to

make a powerful argument that such a causal principle is universal – i.e. that it holds for all

experience whatsoever. If there is a fundamental condition for the possibility of agency that

places constraints on what agents are permitted to do, then such constraints are universal and

apply for all agents. These constraints could rightly be described as objective (since they hold for

all agents whatsoever) and agent-dependent (since they have agency itself as their foundation).

Despite its promise, the weakness of the strategy just outlined can be seen by contrasting it with other agent-dependent evaluative principles. Cultural relativism is the view that what's good is whatever the culture in question takes to be good. If the culture in question approves of killing for fun, then killing for fun is good. If they do not, then it is not good. Of course, this kind of thinking could never furnish a universal principle. If the cultural disagreement is deep enough, nothing can be said to resolve it. If there is agreement on a more basic principle, from which a decision on the question of genocide could be derived, then arguments could begin with that more basic principle. But if there is no more basic principle to appeal to, then the conversation ends. What's more, the relativist will have to admit there is no real disagreement to be had. Since both cultures are doing what's good relative to their culture, both cultures are doing what's good. This sort of agent-dependent principle gives up entirely on the critical force of realism.

However, the constructivist holds that at least some agent-dependent principles are objective. If the constructivist can succeed in showing that there are objective agent-dependent constraints on action, that forbid genocide, then a decisive verdict can be rendered. Using the strategy outlined above, to show that something is good, one must show that taking a particular thing as good is a condition for the possibility of agency in general. It would be question-begging for the constructivist to say that opposing genocide was a condition for the possibility of being a *good* agent. Such a conception of *good* agency would be baseless without an agent-independent notion of goodness. But if the constructivist can show that there are certain things that an agent must approve of in order to be an agent *at all*. It's not hard to see why constructivists see Kantian ethics as providing fruitful insight, and such interpretations have been so influential, that it can be hard not to think of Kant himself as a constructivist. But without presupposing that agents can

just see the superiority of *particular modes of agency*, the constructivist strategy will not be able to yield the moral content required to forbid even the most horrendous actions. The reason why is very simple. No kind of goodness or moral rightness is a precondition for agency.

It's hard to think of a stronger attempt at this than the Kantian argument against the lying promise. According to Korsgaard, when Kant argues that the universalization of the principle of the lying promise destroys the act itself, he shows that the act of lying depends on a rule of action that makes an exception for the liar. The universalization of the principle creates a *practical contradiction* because the liar must make an exception for him or herself. The liar must endorse the principle that most people ought to keep their promises; otherwise the victim wouldn't fall for the lie. The strength of Korsgaard's interpretation is that, if she is right, then the liar endorses the very principle that he or she violates by lying. This gives the Kantian something to say to those who would ask why the liar *has a reason* not to lie. According this view, it conflicts with the liar's own considered view on the subject of whether or not one ought to lie. This is very strong position for a view to be in. Even those who disagree with its verdict actually agree with it! But, unfortunately, they don't. The liar doesn't really support the principle that everyone should tell the truth when it comes to making promises. The liar believes that it's ok for him or her to lie; the self-reflective liar believes the same thing. The only difference is that the self-reflective liar knows that it wouldn't work to lie if everyone did it. The self-reflective liar doesn't care about that, and there's no obvious reason why such a concern is a fundamental precondition of agency. The immoral agent is still an agent. Maybe he or she is a bad agent, and perhaps his or her inconsistency at the universal level tells us something about why this is so. Perhaps there is something inherently superior about consistency in one's willing. But the constructivist refuses

to appeal to an agent-independent conception of badness in order to make this judgment, and it doesn't follow from any plausible claim about the conditions for the possibility of agency. Actual lying agents get along just fine in the actual world. If it's actual, it's possible. So, truthfulness is not a condition for the possibility of agency.

The objection here is that the fundamental psychological, logical, and metaphysical preconditions for agency are insufficient to furnish specific evaluative content. This objection to Kantian ethics has a long history.[56] While it was a brilliant move to consider grounding the rightness in the conditions of agency, if it's going to work, it requires more. Luckily, Kant had the tools available to help meet the challenge. Despite his rationalism and his rejection of speculative metaphysics, Kant was willing to appeal to the phenomena of moral sense in order to defend his value theory.[57] Human beings have moral worth. Human dignity is a result of our rational nature, our ability to think beyond our particular situation and act in accordance with universalizable principles. We are free to set our own ends and recognizing the freedom and dignity of others. These are compelling doctrines that reveal powerful insights into the beauty and value of what is most fundamental to human nature. As a stand-alone moral principle, the formula of humanity has a strong appeal, and the notion that all agents have intrinsic worth should be taken seriously even if it cannot be shown to be a fundamental precondition of agency. If we suppose that a principle of human dignity is an agent-independent objective fact, then Kant's view can be defended on realist grounds. As has already been shown, the task of establishing such a principle is made much easier by positing a reliable capacity of evaluative

---

[56] Classic examples include, **GPR**, 2.3 and Mill, 1863: 5-6, 77.

[57] **G**, 4:393.

perception. Constructivism can succeed if it is supplemented with an agent-dependent objective principle that *we must recognize* the dignity of the good-will. We just see the inherent superiority of a rational universally consistent will over that of the corrupt will of the liar.

1.7. Conclusion

In this chapter, I have attempted to explain evaluative perception and show that, across a wide variety of moral ontologies, evaluative perception has an important role to play in almost any ethical theory that can satisfy the requirements of common sense objectivism.

2.   DOES EVOLUTION UNDERMINE EVALUATIVE PERCEPTION?

2.1.   Introduction

In this chapter, I consider three broad approaches to the evolutionary debunking of moral

knowledge, with an emphasis on the extent to which they succeed in undermining appeals to

evaluative perception. The *selective* debunking of specific varieties of evaluative perception

(emotions like disgust) provides a valuable contribution to our understanding of moral

psychology, but *general* debunking arguments fail to establish the conclusion that evaluative

perception is not a reliable source of moral knowledge. In a further attempt to establish the

conclusion, it makes sense to argue for projectivism by inference to best explanation. I will argue

that this approach also fails to establish the desired conclusion. In the end, evolutionary biology

calls into question certain claims to knowledge by evaluative perception, but it does not

undermine them. This is entirely appropriate. This sort of middle-ground position should be

welcomed by the common-sense objectivist since it justifies subjecting evaluative commitments

to scrutiny without undermining morality in general. I conclude with some observations

concerning how turning our sights toward education and culture may aid in the task of

vindicating certain instances of evaluative perception.

2.2.   Varieties of Evolutionary Debunking

Genealogical debunking arguments seek to undermine claims to knowledge by showing that the

causal origin of the belief makes the belief unlikely to be true. Evolutionary debunking

arguments do this by arguing that the evolutionary history of the human species makes beliefs of

a certain kind unlikely to be true. Evolutionary debunking arguments admit of a wide degree of

domain specificity. At its most general, evolutionary debunking arguments may apply to all beliefs whatsoever.[58] More realistically, they apply to a specific set of beliefs. Of particular interest are those evolutionary debunking arguments that apply *generally* to beliefs about objective moral facts.[59] I refer to this as the general debunking of moral knowledge (see 2.4). I call them general because they apply to all moral beliefs and to distinguish them from *selective* debunking arguments that apply only to specific sources of belief.[60] Here I am concerned with general evolutionary debunking arguments that threaten to prove that the perception of value is not possible and those selective debunking arguments that undermine perceptual claims to knowledge about value.

So how does genealogical debunking work? For every belief that someone has, there is a causal account to be given concerning how they came to have that belief. It's important, then, to distinguish the descriptive and normative conditions of belief formation.[61] The descriptive conditions of belief formation concern those under which a given belief is or is likely to be formed and the normative conditions concern those under which the belief formed is or is likely to be true. Descriptive and normative conditions can overlap, like when you believe that there is mustard in the fridge *because* you opened the fridge and saw a yellow object that looked like a jar of mustard. Since the actual presence of mustard in the fridge is the most plausible

---

[58] Plantinga, 1993. Plantinga is obviously not interested in endorsing this skeptical argument. For him, it is a *reductio* that follows from the combination of atheistic naturalism and evolutionary theory.

[59] Joyce, 2007. Street, 2006.

[60] Kelly, 2011.

[61] In this section, I am using the term 'normativity' exclusively in the epistemic sense, but in section 2.4, I will have to refer to both epistemic and moral normativity. There I will try to avoid confusion by employing phrases like 'epistemic respectability.' To be clear, epistemic normativity is entirely concerned with ways of promoting *true* beliefs and is not to be confused with issues like the ethics of research practices.

explanation for this kind of visual perception, the belief formed on the basis of this kind of visual

perception is likely to be true. But the descriptive and normative conditions of belief formation

are conceptually distinct, and skeptical arguments have often rested on the ways in which the two

can come apart. For example, Joyce uses a science fictional analogy to motivate his evolutionary

debunking of moral knowledge.[62] He imagines a pill that causes whoever takes it to believe that

Napoleon died at Waterloo. He argues that if someone who believes it to be the case that

Napoleon died at Waterloo later discovers that he or she had unknowingly taken one of these

belief-forming pills, then this should raise doubts in the persons mind as to whether the belief is

true. This is reasonable because the descriptive account of how the belief was formed is now

known to be independent of any plausible normative criteria according to which a belief formed

in this way is likely to be true. To be clear, this epistemological argument isn't meant to show

that Napoleon *didn't* die at Waterloo; it just takes away whatever power, or epistemic charge, the

genealogical story of the belief's formation had to support the belief that the belief was true.[63]

The pill-taker would believe the proposition whether there was any good reason to think that it

was true or not. The genealogy of a belief may, therefore, be *vindicating* if appropriate normative

conditions are met, and it may be *debunking* or *undermining* if such conditions are not met.

Genealogical debunking arguments proceed in two stages. One step is to lay out the

genealogy of a belief (or set of beliefs) by articulating and defending a descriptive causal account

---

[62] Joyce, 2007: 179-84.

[63] In chapter two of *The Rationality of Perception*, Siegel (2018) uses the helpful concept of epistemic 'charge' to articulate how justification can be transferred between mental states, allowing her to (temporarily) leave it an open question as to whether phenomenal conservatism is correct in assigning initial credence to currently held beliefs. It seems to me that intuitions about the initial credibility of currently held beliefs, is at least partly based on the assumption that the genealogy of these beliefs is most likely vindicating.

(or a limited range of causal accounts) of its formation. The other step is to show that these descriptive conditions do not fit with the normative conditions appropriate for forming true beliefs. To use a classic example, Sigmund Freud attempted to undermine the belief in the god by rooting this belief in the psychological need for comfort in a heartless and unruly world.[64] On this view, the belief in god results from a deeply felt psychological need that can be satisfied by the belief alone, without the corresponding object actually existing. If this account is correct, then the belief in god is not unlike the belief that Napoleon died at Waterloo in Joyce's science fictional analogy. In both cases, the (supposed) descriptive account of the belief is insensitive to the normative conditions of truthful belief formation. For Freud's argument to succeed, it needs to be the case that his psychological account of spiritual belief is true and that beliefs formed in this way fail to meet the normative conditions of true belief formation. I think that this second of these is quite plausible. In most cases we do not allow that a belief is justified if it feels nice to believe it. But the first of these is much more complicated and beyond the scope of this project.[65] I introduce it here only as a common reference point for understanding the basic strategy of genealogical debunking.

Evolutionary debunking arguments are a subset of genealogical debunking. These arguments rely on facts about evolutionary biology to undermine beliefs in a given domain. In this dissertation, I am concerned with the evolutionary debunking of morality, but evolutionary debunking can apply to any domain of knowledge. The key to these arguments is usually the apparent mismatch between the normative conditions of true belief formation and the conditions

---

[64] Freud, 1927 and 1930.

[65] A sustained engagement with Freud's critique of religion is beyond the scope of this dissertation, but in chapters 3 and 4, I will have some things to say about Jonathan Haidt's more recent comments on the psychology of religious belief and its implications for philosophy.

governing natural selection. Our capacity to acquire knowledge evolved to facilitate survival and reproduction, not for gathering up and storing a bunch of true beliefs. If holding a false belief does just as well by these standards, then there's little reason to expect natural selection to prefer a disposition to hold a true belief. Now, this skeptical argument risks proving too much. If our evolutionary history were to undermine our capacity for belief formation in general, then this would, in turn, undermine our beliefs about our evolutionary history.[66] But when the evolutionary debunking strategy is more localized it can be highly effective. In some domains it's easy to see how the formation of true beliefs could benefit survival and reproduction. Animals have an interest in being able to accurately navigate their surroundings and evade predators. The ability to remember the location of water sources seems to have benefited our ancestors and contributed to their survival by allowing them to chase prey for extended periods without needing to store camel-like quantities of water.[67] It's easy to see how basic reasoning skills could assist in the kind of navigation our ancestors were engaged in, but what quantum physics and differential calculus would have been of little use on the Serengeti. Objective moral value faces a similar problem. We will see that our ancestors benefited from varieties of helping behavior and reputation management. We seem to have evolved to do things for others that we call good. But this is just to say that certain elements of our psychology *feel good.* The history of slavery and brutality in the ancient world is enough to show that survival endures without virtue. If there are objective moral facts, it is hard to see how they come into contact with our evolutionary history.

---

[66] See note 1.

[67] Henrich, 2016: 74.

One strategy for overcoming the evolutionary challenge is called the *companions in guilt* defense.[68] The companions in guilt defense rests on the principle: if an argument entails skeptical conclusions about beliefs that we can confidently take to be true, then the argument ought to be rejected. So, for example, if an evolutionary debunking argument were to undermine basic mathematics, then the argument should be rejected, since we can be confident that basic mathematics is a reliable source of true beliefs. Proponents of evolutionary debunking arguments against beliefs about objective values typically accept this principle.[69] So the strategy of the companions in guilt defense is to show that an evolutionary debunking argument against beliefs about objective value (if accepted) could be generalized to establish skepticism in non-controversial instances as well. This strategy succeeds in rebutting the most extreme forms of evolutionary debunking, but we will see that selective debunking arguments can withstand this criticism, for the very reason that limit the scope to a particular source of belief. What's more, we will see that the most extreme debunking arguments beg the question against moral realism. They are philosophically interesting, but they presuppose a far-reaching philosophical worldview that is not established by our best science. It might be true, but it doesn't follow from what we know about human evolution.

Over the next three sections I will consider the merits of three different approaches to the evolutionary debunking of morality. We have seen that the genealogical debunking strategy involves a two-step process of giving a descriptive account of the origin of a belief (or kind of belief) and then showing that this descriptive account fails to fit with our normative expectations

---

[68] Clarke-Doane, 2012.

[69] Joyce: 2006, 182n5.

concerning how true beliefs ought to be formed. Now we will consider three ways in which the evolutionary debunking of morality can proceed and what they mean for evaluative perception.

2.3.   Selective Evolutionary Debunking

The evolutionary debunking arguments relevant to moral thought can be divided into two broad categories. There are general debunking arguments that seek to undermine the entire domain of moral knowledge. And there are selective debunking arguments that seek to undermine the reliability of a particular source of moral belief. While general debunking arguments are meant to prove that all claims of moral knowledge are unjustified, selective debunking arguments are meant to show that moral knowledge cannot be obtained in some specific way. Selective debunking arguments benefit from greater precision. They are able to draw on empirical research concerning a particular psychological mechanism, without speculating about what a completed science of the mind would look like or making sweeping general claims that risk ignoring the complexities of moral life. They fit well with the tendency toward functionalism in cognitive science and the philosophy of mind, allowing that different mental operations serve different purposes accomplishing them in various ways. It's reasonable to suppose that claims about specific mental operations will be more justified than far-reaching claims about the mind as a whole. One needn't affirm nor deny that the mind is a massive system of modules to acknowledge that cognitive science has yielded profound insights into how the mind works by identifying and studying specific mental functions. Cognitive science can tell us a great deal about how a psychological mechanism may have evolved, what it does, and how it does it. These

insights can then be used to estimate a given function's epistemic reliability in certain kinds of cases.

Indeed, by studying how a cognitive function consistently errs researchers often learn how that mechanism works. Consider a well-known example from visual perception. In the Mueller-Lyer illusion, a line is made to appear longer than another of the same length by adding fins to end of both. The line with fins pointing outward appears longer than the line with fins pointing inward. According to one influential explanation, this is the result of the inappropriate scaling of size constancy.[70] That is, even though both lines are equidistant from the viewer and presented on a two-dimensional plane, the viewer sees them as if one were closer than the other because the fins (seem to) indicate depth. On this interpretation, seeing the line with fins facing outward is like looking at the corner of a room, where the outward facing fins indicate the ceiling and the floor. And seeing the inward facing fins is like looking at the corner of a cube, where the inward facing fins indicate the sides as they project away. Since things that are further away look smaller and things that are closer look bigger, perceptual cues indicating distance fix our relative estimation of size as if the line with outward facing fins were further away. This mental operation doesn't just switch off when the viewer knows that the images are on a flat surface—a fact that seems to lend support to Siegel's rationality of perception thesis.[71] The viewer process the information automatically, altering the phenomenal experience. If this is the correct explanation of what's going on in Mueller-Lyer cases, then we can expect snap judgments of distance based on visual information to be problematic in cases where depth perceptual cues are misleading. But

---

[70] Gregory, 1963.

[71] Siegel, 2018.

this is probably not the only thing going on with the Mueller-Lyer illusion, as it's hard for this explanation to account for similar results in the dumbbell presentation of the illusion—where the fins are replaced with circles.[72] Nevertheless, there is evidence that whatever causes the illusion has to do with a mismatch between the stimulus cues elicited by the whole image (including the fins) and those coming from specific parts (the line itself).[73]

Understanding illusions like these enables us to caution against them. A selective debunking argument identifies a problematic source of information and argues that we shouldn't treat it as a reliable source of knowledge. Certain kinds of perception in certain domains are systematically misleading. We know that the visual system is unreliable in Mueller-Lyer cases partly because we know some things about what the visual system for. In particular, it's *not* for reliably perceiving Mueller-Lyer images! While that's no surprising revelation, the same strategy can be applied to potential sources of knowledge about value in order to discredit theories of evaluative perception. If it can be shown that the origin and purpose of some variety of evaluative perception makes it a systematically misleading guide to objective value, then that we will have to reject its application in those contexts in which it can be expected to be unreliable.

Leon Kass has argued that the feeling of 'repugnance' is sufficient to ground certain moral beliefs. On his view, certain offenses to humanity -- like human cloning -- require no proof to be shown to be immoral. In fact, on his view, sober rational debate about such things risks warping the moral compass. Repugnance, he insists, "is the emotional expression of deep wisdom, beyond reason's power to fully articulate it."[74] It should be clear from what I said in

---

[72] Mundy, 2014.

[73] Mundy, 2014.

[74] Kass and Wilson, 1998: 18.

chapter one that I sympathize with those who would ground evaluative knowledge in a kind of perceptual experience. If there is a reliable capacity for perceiving objective value that is discernible by its phenomenal character can be used to resolve the challenges posed for realists and constructivists in the first chapter. But Kass's appeal to repugnance highlights the inherent difficulties in such a move. Kass's deep wisdom view of repugnance is undermined by the selective debunking of normative appeals to the disgust response. In his 2011 book *Yuck!,* Dan Kelly gives an account of the genealogy and function of disgust.[75] In the final chapter, he situates his skeptical view about the moral justificatory value of the emotion against Kass's *deep wisdom* view and the views of other *disgust skeptics*. According to Kelly, disgust evolved as the result of the functional integration of two distinct cognitive mechanisms, one for parasite avoidance and another for avoiding poisons. These two cognitive mechanisms are now recognized as a distinct emotion which has been co-opted in by cultural evolution to serve certain social purposes.[76] On this view, the feeling of repugnance is not the voice of human dignity crying out for due respect; it's nature's hair-trigger response mechanism for keeping us away from things that can make us sick. As a capacity for evaluative perception, disgust perception is systematically misleading for two reasons. First, sensations promoting poison and parasite avoidance can only be expected to track objective values if objective values are always grounded in the subject's health. They're not. Second, even if health and virtue were identical, these sensations would still be unreliable since these adaptive mechanisms lose little in terms of adaptive benefit by reporting false positives. They have a hair trigger. Rather than crying out in defense of human dignity, disgust

---

[75] Kelly, 2011.

[76] Kelly, 2011: 137-40.

seems better equipped to promote the irrational fear of the unfamiliar and cast out the impure.[77] The strategy of selective debunking begins by identifying the function of a psychological mechanism that might be used to justify evaluative judgments and then shows how it is likely to be systematically unreliable source of knowledge about the good.

Different causal stories may come to the same conclusion. Kelly situates his view relative to what he calls "terror management" views of disgust. The terror management view can also be used to support skepticism about disgust as a reliable capacity for evaluative perception. The terror management view rests on a different genealogy.[78] According to this view, disgust is a distinctly human emotion that facilitates the avoidance of signs that bare symbolic relations to objects of human anxiety – notably, the dreadful (and potentially debilitating) awareness of one's own inevitable death. Human beings seem to be the only creatures who know that they are going to die. According to the terror management view, disgust has the function of turning our minds away from our physical nature. By avoiding reminders of our soft, fleshy bodies, our ancestors were able to devote their cognitive energy to activities that better promoted the flourishing of the species. On this view, if something disgusts us, it probably bears a symbolic resemblance to our bodily nature because it's our bodies that, as they age and get sick or wounded, seem to advertise the fact that we are going to die. Absent the assumption that bodily or organic things are wicked, there is no reason to expect these disgust-perceptions to track moral truth.

---

[77] There's much more that could be said about this. Haidt (2006, 2012) links the disgust response to feelings of sanctity and elevation, which he then associates with group-building and social cooperation. This is strikes me as highly plausible, but if it's true, it has an obvious dark side. Sanctified group-identity grounded in disgust is a recipe for tribalism and scapegoating. The physiological logic of disgust encourages us to violently expel the impurity. Racist propaganda has notoriously invoked notions like infestation, uncleanliness, plague. Perhaps, it is a failure to fully appreciate the dark side of sanctity that has lead Haidt (2012) to extoll the virtues of political conservatism's moral psychology, while others have focused on the psychology of racial bias (Kelly et. al., 2010a, 2010b).

[78] Kelly, 2011: 137-40.

While the terror management view has a nice existentialist ring to it, it struggles to account for instances of continuity with other primates. Other animals express proto-disgust, but the terror management thesis seems to draw a sharp line between self-conscious humans and unthinking beasts.[79] We needn't decide the issue here. The strategy of the selectively debunking forms of evaluative perception is clear. Show that the best available account of the origin and function of a given form of perception is one that is likely to systematically mislead if we were to take such perceptions as evidence of objective goodness or badness.

But there are two important things to note about this strategy. First, selective debunking arguments are limited in their scope. They can only be used to debunk the *particular* psychological mechanism that they are able to explain. Such arguments leave open the general question of whether the mind has *any* faculty for deep wisdom concerning evaluative truth. Second, this approach is inherently fallibilistic and leaves open the possibility that the best available account today might be supplanted by a better account in the future. By limiting oneself to the strategy of selective debunking, one maintains a certain scientific respectability but ends up playing whack-a-mole with the deep wisdom theorist. Perhaps a defender of Kass's view might say that there is subtle difference between disgust and 'repugnance.' The difference may be subtle. Divining a moral verdict from feeling will require the aid of reason and empirical evidence. For example, a recent book argues that we should be skeptical of appeals to empathy. As an alternative, the author proposes we adopt compassion aided by reason.[80] A central task for the moral sense theorist is to distinguish genuine moral feeling from its close imposters. The

---

[79] Kelly, 2011: 44.

[80] Bloom, 2015.

practice of making moral judgments is complex and often takes place on the fly, under the influence of other kinds of emotions, biases, and normative expectations. Even genuine moral emotion can be expected to mislead misinterpreted or applied in the wrong context. But what does this mean for the postulate that there is, for example, a divine sentiment built into the depths of the human heart? Perhaps, that we haven't quite pinned it down yet. If this sounds fuzzy and mystical, it's because it is. No amount of selective debunking can rule out the possibility that *some* deep wisdom view is true. If one is willing to adopt a more speculative metaphysical approach, then options begin to open up. Such a strategy will have to do some philosophical heavy-lifting. One can be forgiven for seeing the endless search for the next 'deep wisdom' theory as the frenzied speculation of those who would flee what they know for the comfort of what they don't. Nevertheless, it is a virtue of the physical sciences that they don't claim to know what they can't; we should be careful not to overstate what selective debunking can prove. As the most scientifically rigorous strongest evolutionary debunking arguments, they are incapable of ruling out a reliable capacity for evaluative perception.

2.4. General Debunking and the Genealogy of Normativity

So, how would a general evolutionary debunking of moral knowledge proceed? We have seen that selective debunking can be effective in illustrating the dangers of drawing moral conclusion from specific emotions, like disgust. Knowing how the psychological mechanism underlying these emotions function makes clear how these may be likely to steer us wrong by failing to track moral truths. So, one strategy for general debunking would be to identify a psychological mechanism underlying *all* moral knowledge and subject it to the same kind of critique. Richard

Joyce comes close to employing this strategy in *The Evolution of Morality.*[81] He argues that all moral knowledge relies on what he calls 'practical clout' and that the psychology of practical clout evolved to facilitate cooperation and pro-sociality among our primate ancestors. So far so good. To follow this up with an analysis of how practical clout may lead to improper snap judgments or prove misleading in the moral search for truth, would surely yield lessons for moral philosophy.[82] But Joyce's goal is more radical. He argues that these observations should support agnosticism about moral facts in general; the project of identifying its evolutionary origin of the moral sense is intended as an epistemological defense of the error theory. In this section, I will show that agnosticism about moral facts doesn't follow from Joyce's genealogy of practical clout. The reason it may seem valid is that it rests on the implicit assumption that projectivism provides the best explanation for the belief in moral facts. I deal with that assumption in the following section.

Before considering Joyce's argument, I'd like to begin by saying what I think these arguments get right. Early in this essay I distinguished between the normative and descriptive conditions of belief formation. Evil demons aside, there are many non-controversial cases the descriptive and causal conditions overlap. You believe that the banana is on the table *because* you saw it there. And you are justified in believing it is there because you saw it there, and that kind of direct visual experience is usually an effective way of forming true beliefs. But since the normative and descriptive stories can—and often do—come apart, it often feels like question-begging to suppose that they overlap in a given case. We tend to assign an initial credibility to

---

[81] Joyce, 2006.

[82] See for example, Pinker, 2011: 622.

the beliefs that we currently hold. That this credibility rests on the assumption that the beliefs were formed in an epistemically respectable way can be seen from the rhetorical effectiveness of genealogical debunking arguments. Simply by becoming aware of the causal origin of a belief calls us to question its legitimacy. The selective debunking of sources of moral belief shows that the assumption of respectability is often wrong. It shows instances in which the normative and descriptive stories *actually do* come apart. The problem becomes more acute when we're talking about the perception of value. It's hard to see how *oughts* and *goods* could be incorporated into a respectable descriptive account without stretching the limits of empirical science.[83] The general evolutionary debunking of moral knowledge tells a naturalistic story about how this might have happened at a very deep level of our moral psychology. The upshot is that I don't think the strategy of seeking out the descriptive foundations of moral psychology has any chance of vindicating common sense moral objectivism without putting its normative cards on the table. The fact that a given moral instinct is a *foundational* feature of our empirical psychology does nothing to vindicate it as source of *moral* knowledge.[84] That we have these foundations is important for understanding how our minds work and how to organize institutions. It's good to know how we might use 'nudges' to get people to save more money and buy certain products.[85] But this is only good because the relevant purposes are good, the consideration of which must go beyond the causal-empirical story to include elements of value theory and moral epistemology.

Joyce begins his debunking argument by identifying what he takes to be the central feature of moral judgment. He calls it "practical clout." Practical clout is the combination of

---

[83] See section 2.5.

[84] Haidt, 2012.

[85] Thaler and Sunstein, 2009.

authority and *inescapability* enjoyed by distinctively moral practical demands.[86] Unlike what

Kant called *councils of prudence*, moral rules place an obligation on agents to obey (**G**, 4:416).

There are two critical steps in Joyce's defense of the innateness of morality. First, he argues that

practical clout is a feature of all genuinely moral judgments; then he argues that the best

explanation for practical clout is that is an evolved moral sense that aided our ancestors in group

cooperation.

The second step rests on what Stephen Jay Gould called a 'just so' story, but Joyce argues

that it is highly plausible—perhaps the best available.[87] According to this account, practical

clout of moral belief was selected because it served as part of *commitment signaling system.*

Signaling a commitment to shared interests plays an important role in group cooperation because

individuals need a degree of confidence that others won't take advantage of their cooperative

efforts.[88] By acting as if moral truths had practical clout—by rigorously upholding moral norms

and severely punishing cheaters—individuals were recognized as trustworthy and reliable. On

this picture, the extremes of moral rigorism and punishment serve a similar function to the

extremes to which members will go to signal group membership (e.g. body modifications, like

circumcision). It also links up with a well-supported view about the psychology of human norms.

Humans have a knack for learning and internalizing norms and learning from the behavior of

those around them.[89] Joyce's thesis adds the practical clout to this innate norm psychology,

insisting that some kinds of norms are imbued with so much importance that they are viewed as

---

[86] Joyce, 2006: 62.

[87] Joyce, 2006: 133-40; Gould, 1978.

[88] Joyce, 2006: 108-23.

[89] Joyce, 2006: 42-43. See also: Henrich, 2016: 185-90; Schmidt et. al., 2012.

inescapable and authoritative. The commitment signaling is, therefore, assured by the presence of a deep-seated moral sense—one good way to convince others that you believe something is to actually believe it.

Evolved commitment signaling is powerful stuff, and it is observable in other parts of the animal kingdom. Early on, Joyce cites an example of reciprocal altruism in the animal kingdom that illustrates the importance commitment signaling for cooperation. Drawing on the work of Robert Trivers, he gives the following description of "cleaning stations" along a coral reef where predators cooperate with potential prey for their mutual advantage:

> Small 'cleaner fish' (or shrimp) indicate their willingness to remove from a large fish its external parasites by approaching the host with a distinctive swimming pattern. The large fish, if it wants cleaning, responds by opening its mouth and gill plates in order to allow the cleaners to go to work. When the host has had enough, it gives a distinctive signal to this effect and the cleaners depart. The host fish could, on any occasion, get a cleaning *and* an easy meal at the end of it… But given that the reef will support only so many groups of cleaners, it is to the large fish's advantage to keep this exchange going.[90]

For reciprocal altruism to work, three conditions must be met: (1) the benefit for each participant must be greater than the cost; (2) free-riders must somehow be denied the benefits of participation; and (3) interaction must take place repeatedly, so that the benefit fair participants outweighs the benefit to a free-rider who just dips in for one round.[91] For creatures like us, an evolved affinity for practical clout steps in where unmediated instinct might fall short. Our ancestors *could* have cheated their neighbors; many of them did. But those who recognized the practical clout of certain norms did better. By their actions, they communicated their

---

[90] Joyce, 2006: 24-25.

[91] For a good discussion of reciprocal altruism that's critical of Joyce's view, see Machery and Mallon, 2009: 24-30.

commitment to the authority and inescapability of certain norms. Through this commitment signaling, they gained opportunity to participate in fitness enhancing cooperative activities. For fish, the instinctive response to the swimming patter is sufficient to direct behavior. Instinctual behavior is more consistent than what humans do. Instinct may run deep, but it doesn't determine human behavior as directly as it seems to with fish. Humans are more likely to cheat. So, cooperation requires additional assurance. Enter guilt, reputation management, and the willingness (eagerness) to punish cheaters. Punishing cheaters and monitoring one's reputation is costly, but the adaptive benefits of cooperation may be greater.[92] One way to build a reputation for holding certain commitments is to hold those commitments. An individual who holds such commitments will be likely to act accordingly in all cases. An individual who only signals commitment when someone's looking is more likely to be caught unaware. The notion of commitment signaling also accounts for the special eagerness people seem to have to be acknowledged for the good that they do. It also explains the sometimes extreme response to betrayals of trust. When someone catches a liar, the intensity of feeling may not stem from the particular offense (which is often quite minor), but concern over the countless other lies that are missed when the liar doesn't get caught. Thanks to the possession of this innate disposition *not* to lie, cheat, individuals gained the cooperation of others and the adaptive advantages that come with it. On this view, what we often call the *conscience* evolved to facilitate reciprocity between human beings by demonstrating an absolute commitment to some basic cooperative norms.[93]

---

[92] Joyce, 2006: 101-105.

[93] Joyce, 2006: 118-123.

If this account is true, then the descriptive genealogical story of the feelings that we associate with a good conscience needn't involve any reference whatsoever to the truth of the evaluative claims it seems to justify. When the intuitionist asserts that lying is wrong and that this is evident from the intuition(s) that arise when you consider lying in the correct way, Joyce can point to our evolutionary history. The intuitionist gets a bad feeling about lying because this sort of phenomenal experience evolved to direct our behavior in fitness enhancing ways. Joyce concludes that we should remain agnostic about moral facts. Our evolutionary history caused our moral belief in a way analogous to his science fictional example of the Napoleon-pill. It makes us highly likely to hold certain beliefs but is entirely insensitive to the truth of those beliefs. The descriptive and normative genealogical stories come apart. If the account is correct, we can expect the moral sense to steer us wrong in certain ways. Hence, Joyce concludes that we should withhold judgment about the truth or falsity of our moral beliefs, a view he calls moral agnosticism.

This radical conclusion doesn't follow. If we suppose that Joyce is right about the genealogy of practical clout, then we should guard against any negative and misleading effects of our tendency treat certain norms as inescapable and authoritative. But this is clearly something that ethicists can do without resigning themselves to moral agnosticism. Consider moral pluralism, on the model of W.D. Ross. On this sort of view, there are self-evident moral principles that always provide prime facie moral reasons to do or not do certain things.[94] These rules are not absolute. In a given instance, the may be overridden by other prime facie reasons. For example, honesty can conflict with the principle of non-harm, as in the famous example

---

[94] Ross, 1930.

associated with Kant (*Right to Lie*, 8:425-30). Telling a small lie to prevent extreme harm may

turn out to be the right in some cases. Making the best determination in a given case requires

good (indeed, fallible) judgment. Given the genealogy of practical clout, we may expect people

to be just likely to overlook the pluralist option as the skeptical option. Practical clout should be

viewed just like we view other biases. Its presence doesn't rule out the possibility of reasoned

debate from coming to the truth on the issue. Indeed, it hasn't stopped some philosophers from

endorsing the error theory and others from embracing pluralism.[95] The presence of bias only

warrants approaching the matter with healthy suspicion. We may justly expect people to be far

too eager to exact punishment, to police inescapable norms at the margins. We might also expect

people to express greater moral outrage at personal betrayals than more substantial harms. We

can expect them to monitor their reputations with vigor, perhaps while failing to tend to more

serious harms—tipping waiters generously while giving little thought to famine relief. Results

like these may well result from a strategy of *selective debunking,* just as they might from the

selectively debunking of emotions like guilt or pity.[96] The genealogy of practical clout warrants a

degree of suspicion concerning the tendency to universalize moral rules and impose strict

punishments. It may even be a good place to start when thinking about intolerance and

intercultural communication. This sort of caution *should inform* our evaluative judgments and

how we interpret certain kinds of evaluative perception.

But Joyce wants skeptical revolution, not reform. To achieve this, he must do more than

provide a genealogy of practical clout. In the chapter where he develops his debunking strategy,

---

[95] Streumer, 2017.

[96] Bloom, 2016.

he supplements his argument by raising some doubts about views that say moral facts are identical to, or supervene upon, natural facts.[97] And it seems to me that, whether he intends it or not, this represents a shift in his strategy. It is not enough to show that some of our deepest feelings about morality evolved to promote fitness. Like a selective debunking argument, Joyce's genealogy of practical clout can only show that the this feature here fails to reliably track moral truth. To carry the *general* debunking strategy forward requires a stronger claim. It must be shown that the output of the process of moral judgment is systematically unreliable. It's difficult, but not impossible, to adjust for biases in the system; we often reject snap-judgments upon reflection. To show that the verdict of moral judgment is systematically misleading is to prove the negative claim that our moral psychology is entirely insensitive to any source of deep moral wisdom. A deep-wisdom theorist who wants to hold the line and insist that the source of moral knowledge is in there somewhere, tangled in physical corruption and buried at the bottom of our tainted hearts, needn't be discouraged by the presence of one more bias. A better way to support the premise is to argue for projectivism by inference best explanation, showing that *whatever* psychological mechanism turns out to underly moral judgment will be unlikely to track moral truth. This is a philosophical strategy that could proceed largely without reference to the results of scientific inquiry. It would be enough to consider only the method of scientific inquiry and the nature of moral judgment. So, in the next section, I will consider the case for projectivism by inference to the best explanation.

---

[97] Joyce, 2006: 190-99.

2.5.   Inference to Best Explanation and the Case for Projectivism

There is a broadly empiricist viewpoint about the relation between objective value and facts that

figures prominently in the work of philosophers like J.L. Mackie and Gilbert Harman.[98] This

view makes it hard to see how moral facts could ever play an explanatory role in the causal order

of things. This view gets something right. Normative principles concern what people ought to do;

descriptive psychological principles concern what people they will do (or probably will do)

under certain circumstances. From this, it seems to follow that we gain no explanatory advantage

by including normative facts in our ontology. From considerations of simplicity, it is reasonable

to exclude from an explanatory account any principles that provide no explanatory benefit. So,

other things being equal, the best explanation of any so-called evaluative perception will not

include any reference to actual moral facts. Our responsiveness to practical clout does not, on

this view, depend on the existence of some objective authority (personal or otherwise) to which

we are responding. Instead, we are psychologically motivated to treat practical clout as

representative of something firm and constant, as if it really were *out there* somehow. To explain

the experience, we may ask if this moral authority is truly objective or just a mere product of our

psychology. Having already accounted for the phenomena by appealing to causal principles,

there is no need to introduce normative ones. In this section, I will argue that this conclusion is

unwarranted. Beginning with the legitimate distinction between causal and normative principles,

the argument trades on their difference while demanding that they meet the same standards. That

is, we should not be surprised that normative principles are unhelpful in causal explanations, but

it doesn't follow that they can have no objectivity or legitimacy of their own. The fault lies with

---

[98] Mackie, 1977; Harman, 1977.

the presuppositions of this empiricist worldview. By applying the principle of simplicity in this way, the case for projectivism begs the question concerning the reality of entities that aren't transparently causal.

Projectivism is the view that (in a given domain) properties that are often taken to belong to an entity or group of entities are produced in the mind of the perceiver. Using this language, we can say, for example, that Locke was a projectivist about secondary qualities, like color and texture, and Joyce is a projectivist about *'practical clout.'* If practical clout is a projection, then the property of being obligatory does not belong to things as they are in themselves. The person who claims to perceive such properties instead *projects* them onto the object. A powerful argument can be made that, from simplicity considerations, inference to the best explanation will always favor projectivism over a theory that posits the reliable perception of objective moral facts.[99] This argument gains additional intuitive strength when we consider the ways in which natural selection has tinkered with our emotions to promote fitness enhancing behaviors. In some cases, human beings certainly do 'project' qualities where there are none – anthropomorphism, for example.[100] One may be tempted to conclude that, *whatever* the best explanation for the capacity for evaluative perception turns out to be, it will appeal to the projection of moral facts, rather than the actual existence of moral facts. I think that this is incorrect, but I should first say a bit more about what projectivism is.

Projectivism is sometimes thought to be open to the following criticism: if the things upon which a given feature is supposedly projected did not have a distinct feature (F-ness), there

---

[99] Rea, 2006.

[100] Joyce, 2006: 123-133.

would be no way of accounting for why the mind projects F-ness onto these things rather than onto others. Whatever it is about the thing that allegedly causes one to project F-ness is what F is in-and-of-itself. Therefore, F-ness isn't projected. It's a feature of the thing-itself. According to this objection, F-ness is whatever it is about a thing that causes us to perceive it as F. This objection fails. While it correctly points out that we need to be able to account for why F-ness is projected on some things rather than others, it is a mistake to say that whatever causes us to project F-ness *really is* that which is meant by F-ness. Consider the projection of emotions. If I say that my infant nephew is grumpy about my sister's absence and it turns out that it was a digestion issue, the fact that he made a distinctively grumpy face is not enough to vindicate my assessment. The actual state of things was just not what I had in mind when I made my assessment. That's the kind of error we're making according to the projectivist view.

As I have already said, I am not engaged in a discussion about the meaning of moral language. The question concerns the existence or non-existence of evaluative facts. That we see things *as if* they were a certain way, suggests a possible counter-explanation for why we seem to see things as good or bad. Maybe we project them. The mind frequently does its work behind the back of the conscious agent, and there is evidence that certain mental functions operate independently to color our experience in (sometimes) misleading ways. The mind is autonomous and productive, working behind the back of the conscious agent – while the agent's attention is elsewhere. The content of our evaluative perception may be colored in misleading ways. If this is something that the mind does regularly, evaluative perception can be explained without reference to objective values or moral facts. Projectivism has the advantage over realism because it can

account for evaluative perception without introducing a novel psychological mechanism. Projecting is something we already know that we do. [101]

The above criticism was supposed to highlight a problem in explaining how why we perceive things as we do. But this is actually where projectivism shines. How do we perceive objective goodness? Is it because objective goodness is an additional quality that human perception has special access to? Or is that we are systematically misled by other observed psychological faculties that tempt us into projecting such features onto the world? Projectivism's success partly stems from hard-won gains of empirical science. But the view enjoys a theoretical advantage as well. Objective values have no obvious role to play in the causal order of the material universe. With the exception of folk-psychology, such principles are rarely invoked in causal explanations. The most plausible cases in which they are concern evaluative perception. Garrett Harman makes short work of such invocations with a classic example. Imagine that you turn into an alley and witness some ruffians pouring gasoline on a cat preparing to light it on fire. The act of animal cruelty causes a feeling of righteous indignation, but Harman argues, "it seems to be completely irrelevant to our explanation whether your intuitive immediate judgment is true or false.[102] So, insofar as there may be psychological explanations available that appeal to causally active features of the mind, brain, body – features that are often already invoked for other reasons – it seems that any explanation of evaluative experience that appeals to the

---

[101] Kelly, 2007: 196. "We have boiled down a large number of the features of cognitive mechanisms underlying our folk psychological capacities to just four properties (albeit high level ones), mainly for easy exposition: implicit and autonomous operation, hair trigger activation, and productive output. These terms describe a complex cluster of mechanisms. They also wear on their sleeve that they describe the functioning of the anthropomorphizer's *mind*, since they characterize the mental operations that give rise to her anthropomorphic tendencies. Indeed, as we shall see, psychologists working in other domains of cognition have posited mechanisms that share many of the same properties."

[102] Harmann, 1977: 7.

existence of moral facts will be multiplying entities beyond necessity. Therefore, the best explanation of evaluative perception will omit any connection to objective values.

Despite the scientific evidence in the premises, this conclusion remains highly speculative. Where selectivity debunking arguments provided a detailed analysis of the ways in which moral emotions can lead us in the wrong direction, the case for projectivism makes a sweeping philosophical claim that veridical objective values never play a role in evaluative perception. Michael Rae has argued that moral anti-realism follows from naturalism, construed as the view that the method of the natural sciences is the only basic source of evidence.[103] The methodologies of the natural sciences are no doubt the best game in town when it comes to the investigation of causal principles. But objective moral principles have testable causal impact, and as such, they are systematically excluded from this kind of investigation. Rather than proving that there are not objective values, Harman's thought experiment, along with Mackie's argument from queerness, might be read as clear limitations of the scientific investigation of morality. If we assume that this empiricist worldview is correct, then we don't need any Darwinian debunking arguments at all. The mere existence of some kind, any kind, of human nature would be enough to establish projectivism. The evolutionary debunking argument aims to show we evolved unaffected by value, but if value can have no effect on anything, this fact becomes trivially true. This is merely to beg the question in favor of the error theorist's favored worldview, sidestepping the more traditional philosophical issues in this area. It is at least equally legitimate to respond to this impasse by looking for theoretical and scientifically respectable ways of broadening this worldview.

---

[103] Rae, 2006: 217.

2.6.   Education and the Vindication of Evaluative Perception

Evaluative perception, as the direct apprehension of moral truth, is a tricky thing. We have seen

that the peculiarity of normative principles makes explanations of it tend toward the mystical;

that a case can be made that we evolved to grant practical clout to such principles; and that the

scientific investigation of an alleged moral sense can bring our attention to its epistemic flaws.

Despite these issues it remains unlikely that this approach to the investigation of human nature

will entirely rule out the possibility of a capacity for objective evaluative perception. But this

would be a sad place for the common-sense objectivist to claim victory. What's needed is a

principled reason for distinguishing reliable instances of evaluative perception from unreliable

ones. Selective debunking shows clearly that we get this wrong sometimes, as when disgust leads

us to confuse parasite exposure with moral turpitude. What principled reason could we have for

not supposing that all evaluative perception is misleading in a comparable way? In this section, I

will suggest some ways in which shifting our focus toward the issue of moral education can help

in this regard.

There are historical examples that help illustrate what a discussion of education can

contribute to our understanding of education. In the *Republic,* Plato elucidates his conception of

a just soul by developing an image of the just state.[104] It's as if we see justice more clearly by

opening the discussion out beyond the individual to consider the broader social context in which

the individual's character is formed. In the *Nicomachean Ethics,* Aristotle argues that only those

who have been "brought up in fine habits" will be able to perceive the difference between right

---

[104] Plato, 1992: 434e.

and wrong.[105] It's hard (perhaps, impossible) to learn virtue if your upbringing doesn't start you on the right path. In early modern philosophy, the emphasis on liberalism and individualism risks obscuring a certain continuity with the ancients on this point. Consider the following passage from Montesquieu's *Spirit of the Laws*:

> Men, rascals when taken one by one, are very honest as a whole; they love morality; and if I were not considering such a serious subject, I would say that this is remarkably clear in the theaters: one is sure to please people by the feelings morality professes, and one is sure to offend them by those it disapproves.[106]

Montesquieu here seems to share the view that there is something about sociality prepares us for virtue and something about the study of culture (their behavior in the theaters) that makes character apparent.

Still, how could education ever serve to bolster the case that a moral worldview is correct or that a moral feeling is veridical? Aristotle's stipulation about culture seems circular when we consider that in one must already be able to tell right from wrong to evaluate the way in which they were raised. With perception as a starting point, one can derive moral principles, but what assures one that these perceptions are apt? The fact that one was raised well. And how does one know that one was raised well? Aristotle doesn't attempt to break out of this circle. He says, instead, that "*we* must begin with what is known to *us*."[107] This is just a limit to our capacity for moral knowledge. In a similar vein, Hegel references the advice of an anonymous Pythagorian on how you can be sure your son is properly brought up. The Pythagorian replied, "when you

---

[105] Aristotle, 1999: 1095b.

[106] Montesquieu, 1989: 5.25.2.

[107] Aristotle, 1999: 1140a-b.

make him *the citizen of a state with good laws*" (GPR: 153). Are we to suppose that a society that

is already corrupt is simply doomed to continue replicating its own false ideology?

Despite these apparent difficulties, the strategy of drawing on culture and education to

vindicate moral perception continues to have proponents today. William Fitzpatrick, for example,

has defended a realist view against the Darwinian debunking strategy by proposing that our

natural capacity for identifying something like a *good* dwelling may have been extended beyond

its original purpose in the same way that our natural capacity for things like counting predators

and estimating distances was extended to theoretical mathematics.[108] He argues that, at best,

evolutionary debunking arguments show that our capacity for moral knowledge was not designed

for tracking objective moral truth—a conclusion consistent with what I have said here. From

there he makes what I consider a very sensible observation (my emphasis):

> The primary lesson to take away from these debates is that as realists develop a
> positive moral epistemology, they are constrained by a need to take seriously the
> fact that natural selection, operating on principles having nothing to do with
> independent moral truths, gave us the raw psychological materials that we are
> now somehow able to *develop, train, and deploy* to arrive at moral knowledge.[109]

Elsewhere, he argues that valuing things like nature and friendship in the right way requires the

cultivation of appropriate "habits and sensibilities."[110] Similarly, John Greco's virtue theoretic

agent reliabilist view reflects an openness to treating moral perception as a learned capacity.[111] If

the evaluative perception is a skill, then, like any other skill, learning to apply it will involve a

process of development or training. Recently, Jason Stanley has put forward an account of

---

[108] Fitzpatrick, 2015: 890-902.

[109] Fitzpatrick, 2015: 902.

[110] Fitzpatrick, 2004: 329-32.

[111] Greco, 2000: 241-47.

propaganda that emphasizes the way in which a false ideology can thwart the reliability of our perceptual beliefs.[112] While his focus is on the negative—the formation and proliferation of *false* ideologies—this naturally suggests the idea of legitimate ideology that would be consistent with the reliable formation of belief.

Certain developments in evolutionary psychology also make me optimistic about this approach. So far, I have emphasized evaluative perception as a direct source of knowledge about objective value, but my description of evaluative perception in the first chapter left open the possibility that the content of evaluative perception might be inferentially loaded. The history of human evolution includes the transmission of cultural information as well as genetic information.[113] Something like Siegel's rationality of perception view might be true.[114] We have also seen that some emotions are noteworthy not just for the way in which they make the subject feel but also for how they communicate information to others. Joyce's view that the moral conscience is a commitment signaling device would make little sense if culture didn't play a role in the genealogy of morality. He draws on the 'gossip hypothesis' and argues that language is a tool for reputation management and reciprocity. On this view, language didn't evolve for primarily descriptive purposes; it has to do with communicating evaluative information to and about those with whom we must cooperate.[115] Even emotions like disgust, which at first appear automatic and subjective, turn out to be cognitive and communicative. Kelly draws on the work

---

[112] Stanley, 2016.

[113]See, Boyd and Richardson, 2005: 191-236 and Henrich, 2016. As Henrich puts it, purely genetic theories cannot: "explain our ecological dominance or our species' uniqueness without first recognizing the intense reliance on we have on a large body of locally adaptive, culturally transmitted information that no single individual, or even group, is smart enough to figure out in a lifetime" (12).

[114] See section 1.2.

[115] Joyce, 2006: 88-92.

of Boyd and Richardson to explain the social function of disgust in terms of what he calls its

"sentimental signaling system."[116] When someone is disgusted, we can (often) literally see it on

their face. This warns us to steer clear of whatever's causing the disgust response. Disgust is

cognitively rich in another way too. Rozin et. al. observe that "disgust is linked to sympathetic

magical thinking" and that disgusting objects can "contaminate the things they touch."[117] The

logic of contamination allows things that are not disgusting to induce disgust, merely by virtue of

certain relations. All of this suggests that evaluative perception is often inferentially loaded. If

our emotions are at all sensitive to this logic, then there is a role for reason and the understanding

in helping to calibrate our evaluative responses and bring them into line with objective moral

judgments.

Paul Bloom's recent book on empathy has an admirably straightforward title: *Against*

Empathy: The Case for Rational Compassion. In it he argues quite convincingly that

unrestrained appeals to empathy often lead to bad moral judgments. He argues that empathy is

spotlight emotion that places too much importance on the feelings of those right in front of us,

while failing to appreciate the concerns of those out of view.[118] But it seems a bit unfair to

compare one moral emotion on its own (empathy) to another moral emotion (compassion) aided

by reason. In truth, moral feeling will need to invoke reason at a variety of levels. Whether

empathy or compassion motivates me to aid the victims of natural disasters, I will need reason

and experience to make a judgment of how best to do so. If empathy or compassion makes me

initially aware of the prime facie badness of harm, I will need reason and experience to discern

---

[116] Kelly, 2011: Ch 11.

[117] Rozin, 1999: 332. See also, Joyce, 2006: 96-97 and Kelly, 2011: 19-20.

[118] Bloom, 2015.

whether some other principle should take precedence in a given instance. Reason and experience may be needed to disentangle genuine cases of empathy or compassion from imposters, like tribalism. Perhaps pure empathy is a kind of divine revelation, telling us to care deeply for humanity. But this revelation is so thoroughly contaminated by earthly considerations, so obscured by an individual's limited perspective, that the only lesson we can draw from it is that vague call of a thoughtful conscience—a seemingly trite tautology, like, 'it's good to be good to people.' If there is such a revelation at the heart of our moral experience it makes sense that culture would play a role in drawing it out. By creating opportunities for individuals to come face to face with this kind of fundamental experience, cultural practices can play a role in making us more attentive to the good. One of the good things about going to church, or on a long hike, or to yoga class, is that these things tend to bring peoples minds back to what they truly feel is important. These impractical settings prepare us to receive information that we might resist or misuse if it came at us on the fly. People can be feel the truth of messages delivered from the pulpit that they'd never assent to in a debate at the bar after work. I see nothing in Bloom's book to suggest why empathy can't be a source of genuine moral insight in this sort of context.

The problem of circularity noted above, seems to stem from a difficulty we have in making sense of the *immediacy* of immediate apprehension. Immediacy seems to suggest that one either has it or they don't. So, there's tendency to conflate all immediate apprehension with personal preference. You either like Miller Light or you don't. When it comes to deeper moral commitments it starts to look like a question whether you've got the stomach to deny moral facts or if you need to seek refuge in some form of deep-wisdom theory. The difficulty we have

denying the error theory becomes a case in its favor.[119] I think that this impression depends on a superficial genealogy of moral experience. There are two reasons why bringing education into the conversation may remedy this. (1) it's just easier explain how someone was prepared to have an immediate experience of the good than to explain the actual experience, which may be unintelligible. This is why Socrates had to magnify justice in the *Republic* by making it about the city and why he expresses embarrassment at his analogy of the sun in Book VI.[120] (2) The genealogy of an experience, which includes education as a kind of preparation, gives us something concrete to consider when evaluating whether they came to the experience in the right way. I will say more on this in a moment, but this is what I think makes genealogical debunking arguments so compelling. Finally, (3) the account of the build-up to a direct experience gives us something to imitate—actual imitation or by an act of mind—to see if we can have the same immediate experience for ourselves. This last one is often impractical, but the point of all three is that education and genealogy give you something to get your head around, something to consider when assessing how seriously to take claims about the immediate perception of value. The genealogy of immediate apprehension may be the only thing we can consider in evaluating whether the immediate verdict the perception is justified. Education, as a kind of preparation, is the aspect of that genealogy most likely to vindicate a given instance of perception.

I can think of three ways in which this aspect of the genealogy can help to assess the reliability of an immediate apprehension: considerations of reliability; considerations of consequence; and considerations of autonomy and alienation. *Considerations of reliability*

---

[119] Streumer, 2018.

[120] Plato, 1992: 506d-e.

concern the way in which one comes to have an experience of immediate appreciation and what implications it may have for how seriously we should take the experience. For a negative example, consider road rage. After being cut off in traffic, one feels anger and hatred toward the other motorists. This same person may express desire to see others harmed. We shouldn't take this feeling seriously, because the context in which in which it was formed makes the individual an unreliable judge at that moment. It's likely that, in a more amenable environment, this same person would make moral judgments entirely at odds with the road rage feeling. Education and culture can play a role in creating possibilities for more reliable introspection. I've already mentioned churches. Cultural practices like funerals allow those who've just lost a loved one to face that hard reality in the company of loved ones and spiritual advisors who've see death more regularly than the typical person is likely to. Family members are invited to, "say a few words." Customs like these have developed over generations and have endured, in part, because they help to satisfy the deeply felt human need to grieve at the loss of a loved one. It is no refutation to point out that not everyone finds closure or solace in a funeral. It's enough to point out that it provides a better environment in which to begin to come to terms with pain and loss than what one might find in some more everyday contexts, like driving on the freeway. The fact that, in the past, people found satisfaction in a practice like this provides some evidence that one might find satisfaction in it now. Of course, they also, might not. The important thing here is that the genealogy of direct moral feeling can be judged in terms of whether the context created for it is likely to make it more or less reliable.

Second, there are *considerations of consequence*. If a profound experience motivates a person to make drastic changes in their life, one hopes that these changes would be for the better.

It's in considerations of consequence that Bloom's critique of empathy is most persuasive. When empathy leads large numbers of people to ship toys and letters to disaster victims, the resulting stockpile does more harm than good.[121] If someone claims to be directly aware than *more* empathy is *always* for the better, it's to remind them of cases where empathy action can get in the way. In the next chapter we will see that Rousseau's attitude toward health was, well, unhealthy. Culture and education can mediate social friction, but it can also lead to greater problems. If a cultural practice becomes destructive or obsolete, this may over time lead to its being abolished. We shouldn't be blind to the ways destructive and contradictory social practices can endure, but an institution's ability to endure can sometimes provide evidence that it serves an important purpose.

Finally, there are *considerations of autonomy and alienation*. This may sound out of place, but each of the three thinkers put freedom and alienation at the center of moral reflection. They have profound things to say about the lived-experience of a creature capable of self-determination. Arguably, the lived-experience at the center of the distinction between normative and causal principles is a good place to start if we're looking for the ground of moral feeling. Rousseau describes the fundamental experience of freedom as being divided against oneself. Freedom has the sense of the possibility of being different from what you actually are. 'I could have been a better basketball player.' 'I could have said something different.' There is a temptation to compare oneself with others and with what one might have been. 'I wish I'd said something different.' One can even say, 'that wasn't *really* me.' Rousseau's notion of conscience is complicated by describing it first as something *solitary* and complete and then as something

---

[121] Bloom, 2016.

born of one's fellows. The deep duplicity of human nature is tied culture. The process by which we have hope of returning to ourselves is education. Pinkard describes Kant's paradox of autonomy.[122] Autonomy involves being bound by a law that you give yourself. The fundamental experience of autonomy as of a divided self is arguably the central theme of Hegel's famous master-slave dialectic. If the inner tension that characterizes autonomy is manifested socially, say, as a struggle for recognition, cultural evolution will be partly formed by something sensitive to a deep feature of the human experience. Ideally, this would steer us toward customs and rituals that satisfied each individual's need to be acknowledge and respect as they are in themselves. Of course, this ideal is never really achieved. But I will argue that by considering customs and educational practices in light of this ideal we gain valuable insights and appreciation for certain social practices and a helpful criterion for determining whether a social practice may help facilitate reliable moral perception.

Slavery runs contrary to all three considerations. It is fundamentally inattentive to the *autonomy* of the enslaved person. This fundamental inconsistency with human nature has the *consequence* of making it a deleterious and unstable system. And, finally, the persistence of slavery creates a context in which evaluative perception is *unreliable*, since the oppressed are systematically dehumanized and oppressors unjustly exalted. Instances of evaluative perception are usually culturally embedded and subject to environmental influences; these considerations allow us reflect on evaluative perception in context. What we get is not a foundationalist moral epistemology but, rather, a way of refocusing the conversation in a productive way.

---

[122] Pinkard, 2002.

2.7.   Conclusion

In this chapter, I explained how genealogical debunking arguments work and considered the

subset known as the evolutionary debunking of moral knowledge. After considering three

approaches to the evolutionary debunking of moral knowledge, it has been argued that these

cannot rule out the possibility of a deep-wisdom theory of moral perception. The more limited

'selective' debunking arguments have been shown to provide important insights into moral

psychology, but the attempt to extend the debunking strategy to morality in general fails. Finally,

we have considered some ways in which invoking culture and education may have the potential

to provide a limited vindication of moral perception.

3. CONSCIENCE AND SOCIAL TRANSFORMATION

3.1. Introduction

In this chapter I will explain and defend what I call Rousseau's 'egalitarian perfectionism' and it's two-fold structure of vindication. This view can withstand the genealogical challenged outlined in the previous chapter. In section 3.2, I will dispel some concerns that stem from the popular misinterpretation of Rousseau as a so-called primitivist. Then I outline my interpretation of Rousseau as an egalitarian perfectionist. After that, I consider the view in light of evolutionary debunking and moral dumbfounding, to illustrate how genealogy can help vindicate evaluative perception.

3.2. On Hating Doctors

In his book on education, Rousseau expresses contempt for medicine (Emile, 55, 82). Extreme comments like this call to mind an obvious objection. If everything natural is good, then there are no natural evils. But there are natural evils, e.g. disease. I have no interest in defending these comments about doctors. He was obviously wrong if his view would imply that children would better die than be vaccinated for malaria. I don't think that the core of his view implies this, but he probably would have. He came remarkably close when he suggests to Voltaire that the victims of the Lisbon earthquakes were being punished by God for building unnaturally tall six story buildings (*Oeuvres IV,* 1061). He was dead wrong, and it wasn't just and an off-hand remark. It is central to his view that *everything* god creates is good (Emile, 37, 92). In interpreting Rousseau, we have no choice but to take Rousseau's naturalism seriously.

So it's not surprising that when Rousseau gets mentioned by cognitive scientists and psychologists, it's usually to quickly dismiss his view as a cautionary example: Beware the myth of the noble savage![123] These authors provide good reasons for rejecting the view that our instincts are always virtuous and that it would have been all peace and love, had we only remained in the so-called state of nature.[124] As we have seen, the problem with this *primitivist* interpretation is not that lacks support in Rousseau's text. The problem is that it can't account for other things that Rousseau also said. This gives rise to another objection. Rousseau's view is incoherent. The early Rousseau, so the story goes, was an individualist who saw society as a fundamentally corrupting influence, but as he developed a proto-fascist view of the general will —an Orwellian nightmare in which the individual is paradoxically "forced to be free" (*Social Contract,* 1.7.8)[125] In the following section it will become clear that the charge of contradiction fails to pay sufficient attention to Rousseau's distinction between natural and civil freedom. In the rest of this section, I will show that more philosophically engaged comments about natural man reflect a moral philosophically sophisticated view than the textbook primitivist interpretation would imply.

What I'm calling primitivism implies three closely related theses. The first two are descriptive. They amount to the idea that (1) humanity's original nature is good and (2) social life corrupts this nature. From this, a normative thesis is seems to follow. Namely, (3) being good involves a replication of or return to that original state. As an interpretation of Rousseau, this

---

[123] In this section I will focus on Pinker, 2011. His treatment of Rousseau is more sustained. Examples can also be found in: Haidt, 2014: 381n5; Prinz, 2012: ix-x.

[124] Pinker, 2011: 35-36, 432-33.

[125] Russell, 2004: 675-701.

view is supposedly what Rousseau had in mind when he wrote his *Discourse on the Origin of Equality Among Men*—hereafter, *The Second Discourse.* In this text we find Rousseau developing his theory of alienation and providing a conjectural history of human social organization. To be sure, this history of alienation reads a lot like a fall from grace. But if Rousseau was committed to (1) and (2) it doesn't follow that he was committed to (3), and while he was, in certain sense, committed to (1) and (2) we must be careful about what this commitment really amounts to.

In his recent book on the history of violence, Steven Pinker makes the case that we are less violent than our ancestors, and he targets Rousseau for romanticizing the state of nature. Pinker contrasts Rousseau's view with Hobbes's, associating Rousseau with contemporary standards of political correctness.[126] Pinker's concern here is to do away with the false dichotomy between "the nasty brute of textbook Hobbes and the noble savage of textbook Rousseau" and insist that "there must be a grain of truth in conceptions of the human mind that grant it more than one part."[127] I suppose the word 'textbook' gets him off the hook. Maybe he didn't need to bother quoting Rousseau at all, but since he did, I have to point out that the quotation has holes you could drive a darn boat-show through:

> Hobbes's opposite number was the Swiss-born philosopher Jean-Jacques Rousseau (1712-78) who opined that "nothing can be more gentle than [man] in his primitive state… The example of the savages… seems to confirm that mankind was formed ever to remain in it., …and that all ulterior improvements have been so many steps… towards the decrepitude of the species."[128]

---

[126] Pinker, 2011: 35-36.

[127] Pinker, 2011: 482-83.

[128] Pinker, 2011:35-36. Original in *Second Discourse,* Part II, Paragraphs 17-18.

This is not the only passage from the *Second Discourse* that supports the primitivist interpretation, and I give Pinker credit for not omitting the word 'seems.' But when this passage in particular is taken in context, it supports my suggestion that Rousseau is working through something much philosophically richer than the primitivist thesis.

A major thesis of this chapter is that some of the apparent contradictions in Rousseau's thought stem from a serious engagement a fundamental duality in human nature. Rousseau viewed the human mind as capable of action and self-reflection. And these features complicated his account of human nature. The primitivist thesis depends on the view that humanity has an original nature and that this original nature is good. Rousseau agrees. The *Second Discourse* tells a conjectural history of how we got from that nature to where we are now. At the center of this story is a theory of *alienation*. Alienation is a complex concept, but at its core it refers to the way in which the human capacity for action and self-reflection pulls us away from our original nature and sets us in a fundamentally different relation to nature than other creatures. The history of human beings is multilayered, because the way that people *view themselves* and what they are doing influences what they do. The levels of analysis begin to compound on each other, and this can create multiple legitimate stories of why certain events took place. The physiological analysis of human nature exists alongside—and sometimes intermingles with—the stories that we tell ourselves about why we do what we do. In the passage quoted by Pinker, Rousseau is trying to sketch the history of inequality at a deep psychological level by focusing on pride, which he views as an emotion that results from human self-reflection at an early stage. "The example of the savages" referred to in the passage refers to a period in human development that Rousseau calls "the genuine youth of the world" and "the happiest and longest lasting epoch."

This stage is just as far from "the stupidity of the brutes [as] the fatal enlightenment of civil man" (**DOI**, 2.15-18). At this stage, pride has already begun to have a corrupting influence; it cannot be taken as an example of a time when humans were perfectly good.

We must also keep in mind that this history is conjectural. This is necessary because of the object of inquiry. For Rousseau, human nature is somewhat unstable because it depends so heavily on self-conception. The way a person is can be deeply transformed by the way in which they understand themselves. In the preface to the *Second Discourse* he tells us, "in a sense it is by dint of studying man that we have made it impossible to know him." We can't how people originally were by studying how they behave now. This may be a dig at Hobbes, who at one point in *The Leviathan* infers that the state of nature was violent from the fact the people still lock up their property.[129] To the extent that this is a fair interpretation of Hobbes's remark (and I'm not saying that it is), he is engaged in unscientific speculation. Pinker makes a good point when he says that the debate would have benefitted from more and better empirical data.[130] Nevertheless, Rousseau's concept of alienation and the transformation human nature gave him philosophical reasons for acknowledging that inquiry regarding the origins of society will always involve an element of speculation.

At this point, it may sound like my defense of Rousseau proves too much. If human nature undergoes a historical transformation, then how can we talk about corruption at all? For all we know, we were scoundrels from the beginning. Maybe we've even gotten better. Rousseau's assumption of the goodness of natural man is not founded on his conjectural

---

[129] Hobbes, 1994: 1.13.10.

[130] Pinker, 2011: 36.

anthropology. It does not imply the superiority of indigenous cultures nor does it justify any effort to go back to a simpler time. But his insistence that what is natural is *good* is fundamentally theological. To this extent, it is fair to describe his approach as unscientific, but we will see in the next section that there is a crucial difference between the way in which the natural world is good and human moral virtue.

And we should be careful not to take the criticism Rousseau's religion too far. Joyce, for example, argues that there is a problem with the thesis that morality has an entirely cultural origin. He targets another famous passage from the *Second Discourse*:

> The first man who, having enclosed a piece of ground, to whom it occurred to say *this is mine*, and found people sufficiently simple to believe him, was the true founder of civil society. How many wars, crimes murders, how many miseries and horrors mankind would have been spared by him who, pulling up the stakes or filling in the ditch, had cried out to his kind: beware of listening to this imposter; you are lost if you forget that the fruits are everyone's and the earth no one's (**DOI,** 2.1).

Joyce worries that Rousseau means to imply that the very notion of mine-*ness* is culturally constructed. If so, he argues, then the proclamation, "this is mine," would make no sense to them. If they have no notion of morality at all, then there's no way of explaining it to them, since no appeal to ownership, rights, or fair or just treatment would have meaning for them. "The others may be just as smart as you when it comes to observation and inference, but if they don't already 'get' moral thinking, then it is a mystery what might be said… in order to teach them."[131]

This gets to a common misconception about Rousseau's view of human nature. While we can't *know* what we were like before culture, it doesn't follow that *everything* about our nature originates in culture. As Joyce recognizes, the problem with the ontogenesis of morality is only a

---

[131] Joyce, 2006: 183.

problem if Rousseau completely rejects notion that we are endowed with innate moral concepts. But Rousseau sees the moral conscience as endowed by God. Furthermore, there is nothing I can think of that precludes that mind has innate capacities and "more than one part," as Pinker suggests. Rousseau's view allows for a lot of flexibility on this issue. It's true that he had a dearth of data to work with. Ironically, his Christianity leads him to say some things about human nature that fit well with talk of 'reverse engineering' engineering the we hear from cognitive scientists, like Pinker, today.[132] Rousseau insists that we are endowed with instincts that make us better able to navigate our environment; he just thinks we got these from God. His talk of instincts is vague and theological: "The earth's products provided [natural man] with all necessary support, instinct moved him to use them" (*Second Discourse,* II.2).[133] We needn't assume that Rousseau would have rejected the modularity thesis.

Rousseau doesn't see the mind as a blank slate. His view may be helpfully compared to that of Jesse Prinz. Prinz rightly sees this metaphor as something of a straw man, if for no other reason than that "blank slates don't learn anything." Debates between 'naturists' and 'nurturists' concerning a particular innate ability tend to involve disagreements about the degree of specificity with which the natural basis for an ability operates. On Prinz's view, the disagreement in each case rests on a continuum, where one extreme includes faculties like color detection and the other includes social practices like games. Color detection is a trait that has been selected for because it enhances fitness. The ability to play baseball was not. The ability to throw seems to be

---

[132] Pinker, 1997.

[133] I suspect that part of Rousseau's reputation as an anti-nativist stems from the fact that many of his actual scientific claims about human nature are indefensible. They range from outdated to the ridiculous. His comments on women tend to reify sexist prejudices with mere conjecture. Some good examples can be found here: *Emile*, 44-47, 78-9, 82, 131, 361-62, 365.

an evolved human trait. But the game of baseball rests on a system of rules and conventions that could have been very different. [134] One of my major concerns in this dissertation is the way in which human beings, through cultural practices and traditions, take control of their activity and direct innate capacities in toward various ends. If the rules of a game like baseball had been different, then the practices and behaviors surrounding the game would likely be different too. If home runs counted as foul balls, then there would be little incentive for players to beef-up and try to knock it out of the park. What we do reflects our values, but it also turns back around us and helps determine our values by situating our interests and directing our attention. Social institutions can set incentives explicitly through threatening punishments and promising incentives. Culture may also determine our values in subtle ways. Cultural norms and ideology can determine what counts as odd and what stands out in a given situation. One reason that businesses advertise to help make sure that when the consumer becomes aware of the relevant need or desire the advertiser's brand is the first thing that comes to their mind. Rousseau's view, like Prinz's, suggest that we should first look to such factors when explaining human behavior. If culture provides the explanatory mechanism, it is to the roots of culture that we may want to turn in order to explain the more interesting aspects of the ontogenesis of morality—even if we accept Joyce's point that some innate capacity for moral comprehension is required.

In this section we have seen that Rousseau's naturalism sometimes manifests itself in unfortunate comments about doctors and natural disasters, but at its this is a theological view about the relation of God to creation. The view does not imply the normative thesis that we should copy the lifestyle of indigenous peoples. It is compatible with nativism in psychology and

---

[134] Prinz, 2012: 6-7. On throwing, see Henrich, 2016: 290-91. For an interesting example of cultural variability, see the way baseball (or, *Pesäpallo*) is played in Finland.

the modularity of mind, but it also puts culture and self-reflection at the center of any account of human nature. Rousseau's own account begins with some philosophical reflections on how creatures who (by nature) reason, feel, and act come to prefer some things over others and act in ways that reflect these preferences. He develops a perfectionist account of the good life founded on a deep-wisdom theory of the moral conscience and a structural account of pride and alienation. And it is to this account that we now turn.

3.3.   Rousseau's Egalitarian Perfectionism

I call Rousseau's view 'perfectionist' because it calls individuals to achieve their highest good. To perfect oneself is to complete a project that nature could only begin. I call this view 'perfectionist' because this completion involves a return to what is most common. All human beings naturally sense and *love* the good. The notion of moral superiority is, thus, distorted and corrupt. To perfect oneself is not to elevate oneself above others but to knowingly reconnect with what was always already present through sensation. Pride is a symptom of a deep psychological potential for corruption.[135] This potential is an inevitable result of our distinctive psychology, without which we could not possibly *know* the good. While nature (excluding humans) is unified and perfect, human nature involves a fundamental rupture. To become whole again, we must knowingly return to those deepest sentiments that are first and most purely available to everyone. So, even if it is a rare achievement, our highest good comes only with the knowledge that we are no better than anyone else, that everyone alike is tethered to the good by moral sentiment.

---

[135] I will use 'pride' to refer to what Rousseau calls *'amour propre'* and 'self-love' to refer to *'amour de soi.'*

Rousseau is unwavering in his commitment to the following two principles. First, (1) everything natural is good. Second, (2) human beings corrupt (Emile, 37). But there is an obvious tension. If (1) is true, then (2) seems to be false, since human beings are, or at least seem to be, natural entities. So, what is it about these apparently natural beings that causes them to corrupt nature? It's easy to get lost in Rousseau's rhetoric, but if we keep this question in focus, his solution turns out to be remarkably consistent, given the complexity of his topic. For Rousseau, human beings have an innate capacity for evaluative perception that he calls "conscience." In *Emile,* the Savoyard Vicar provides an admirably clear definition:

> There is in the depths of the human souls, then, an innate principle of justice and virtue according to which, in spite of our own maxims, we judge our actions and those of others as good or bad. It is this principle that I give the name conscience (**EE**, 289).

And a few paragraphs later he asserts with equal confidence:

> It is from the moral system formed by this double relation to oneself and one's fellows that the impulse of conscience is born (**EE**, 290/**OC** IV, 600).

The conscience is innate, but the impulse of conscience is born from our relation to others. One might want to resolve the apparent contradiction by noting that the first passage refers to a 'principle,' and the second refers to a separate 'impulse.' But this is hard to square with Rousseau's sentimentalist characterization of the conscience is a feeling that makes us love the good long before we know it (**EE**, 67, 235, 286). Paradoxically, it is better to allow that the conscience has a dual origin, in a sense both innate and born of social interaction.

This fits well with some of Rousseau's other paradox's. We have already seen that human beings should be both naturally good and inherently corrupt. And in *The Social Contract* he describes a two-fold relation to the state, according to which one is both sovereign and subject

(**SC**, 1.7.1). And, of course, there is his distinction between natural man, citizen, and bourgeois. The latter is a deformed mis-mash of the other two. One can be a citizen or a man, but not both. According to Rousseau, we are born with the use of our senses, which he later tells us are passive and add nothing that is not already given in nature (**EE**, 39, 270). Natural man is a creature of feeling, unified and complete. The citizen is a member of a greater whole. "A citizen was neither Caius nor Lucius; he was a Roman" (**EE**, 40). The citizen is the one whose conscience has been transformed completely by their relationship to the state. Conscience in the first sense—as an innate principle—belongs to every person by their nature, but it is fundamentally transformed through one's relation to others. One way to express this is to say that conscience belongs to us all implicitly as an innate principle, but it is only explicitly realized through social life.

So culture transforms us at a fundamental level. Since this transformation can go well or badly, we risk turning us away from our innate access to the good and becoming corrupt. But what does 'culture' mean here? The citizen is obviously shaped by culture, but in the discussion of conscience, Rousseau refers to a 'double relation' and one's 'fellows.' Again, recall Rousseau's fundamental principle: "everything is good as it leaves the hands of the Author of things; everything degenerates in the hands of man" (**EE**, 37). God made everything perfectly. Nothing could be done to make it better. It can, however, be made worse. And that's where we come in. We must read the transformation of conscience in terms of a deep psychological potential for corruption. Rousseau's moral psychology is grounded in the *imago dei* doctrine and informed by a conception of original sin. From birth, human beings have a capacity of passive sensation that allows us to *sense* and love the good. In this way, we are naturally perfect. But, this passivity is insufficient for knowledge. Instinct would have been enough to facilitate sexual

reproduction, nursing, hunting, and bee-like cooperation. But God wanted us to experience the good more like how he does, by knowing it. For Rousseau, the "supreme enjoyment" for which we were created is to be happy and to know that we deserve to be happy (**EE**, 281). To facilitate this, God also gave us the active faculties of *will* and *judgment*. In so doing, he gave us a role in a kind of second creation, wherein we transform ourselves and shape the world around us. But, since creation is already perfect, our changes only make things worse. And since everything as already good the goal of education, of our self-transformation, is not to improve upon what God created but to approach its original goodness in a knowing way.

By giving us the active faculties of judgment and will, God made us moral; he gave us the "right to virtue" (**EE**, 281). But these active faculties also created a fundamental rupture in our being. For Rousseau judgment is a form of comparison. Sensation puts us in direct contact with things "such as they are in nature" (Emile, 270). But judgment involves a kind of mental juxtaposition:

> By comparison, I move [objects in sensation], I transport them, and so to speak, I superimpose them on one another to pronounce upon their difference or their likeness and generally on all their relations (**EE**, 270).

When we make judgments, we tamper with what's given, and in so doing, we degenerate the perfection that we were given. We cannot make judgements without corrupting God's perfect creation, but we can't achieve our highest good without making judgments. So, while God wants us to know the good, we can't know it in quite the same way he does:

> Man is intelligent when he reasons, and the supreme intelligence does not need to reason. For it there are neither premises nor conclusions; there are not even propositions. It is purely intuitive; it sees equally everything which is and everything that ought to be (**EE**, 285).

God's intellectual intuition allows him to know the good immediately, without going to work on it. God's intelligence is a pure unity. It has no need for judgment, comparison, or inference. God knows can love purely because he knows he is good without comparison. He needn't check his goodness against some ideal; he intuits it purely. Humanity occupies the complicated middle-position between the pure unity of sensation enjoyed by animals and the unity of absolute intellect enjoyed by God.

This deep psychological potential for corruption leads to a characteristic way in which human beings can be expected cause things to degenerate. In the *Letters on Botany*, Rousseau cautions against studying disfigured flowers. "When you find doubles, don't get caught up with studying them. They are disfigured, dressed in our fashion, if you will" (**OC** IV, 1156). To be duplicated—one might say, 'duplicitous'—is to share something in common with us and entirely inappropriate for the rest of creation. In another of these letters, he attributes the duplication to human causes and suggests that one avoid studying flowers as they are found in the garden:

> Man has distorted [*denaturer*] many things in order to better convert them to his use. In so doing, he is not to be blamed, but it is nevertheless true that he often disfigures [*defigurer*] them. Mistakenly, he thinks that he is truly studying nature, when he is studying works of his own hands. Mostly, this error takes place in civil society, but it's the same in gardens. These double flowers—that one so admires in the flower beds—are monsters. They are devoid of the capacity nature has given all organized beings, the capacity for producing its own kind [*leur semblable*]. It's roughly the same with the fruit trees… So, to know the pear and the apple of nature, you must seek them out not in the gardens but in the forests. The flesh is not so large and succulent, but the seeds ripen better and multiply more; the trees are infinitely larger and more vigorous. But this point is taking me too far afield. Let's return to our gardens. (**OC,** IV, 1188).

In flowers and trees, this duplication is a monstrosity. For us, it is a necessary evil—a precondition for the achievement of our highest good. In order to know that one is worthy of

happiness, one must be capable of making a judgment about one's worth. This judgement is comparison. To judge that I am good I must first have a sense of what I am. Then, I must judge that what I am is good. God is conscious of his own goodness without making any judgments. Divine self-love never degenerates into pride because knowledge of God's goodness does not require comparative judgement. Knowledge of his own goodness belongs to God as a kind of immediate appreciation. How this can be *knowledge* is inevitably a mystery to us. We achieve self-awareness through a kind of comparison. We compare ourselves to others or to what we might have been. In this way, we remove ourselves from our original unity and juxtapose ourselves against someone or something else. We construct and ideal out of the good by applying our comparative judgment to our original sense of the good. But this ideal is at best approximation of what can only be known be known to God.

From this deep inner separation results the distinctively human emotions of pride, envy, and pity. Self-love (*amour de soi*) is an essential characteristic of any self-aware being who loves the good. A self-aware being loves itself and wants what is good for itself. Even God has self-love. But human beings can't have self-awareness without judgment. We develop a sense of ourselves and juxtapose it to the ideal. Thus, pride (*amour propre*) is a distinctly human emotion that combines self-love and the comparative faculty of judgment. Human beings naturally love the good. From this, it follows that when a person compares one thing to another thing, they will prefer the better. But as self-awareness develops, self-love begins to win priority over the love of the good. The subject gains satisfaction in winning the comparison. This satisfaction is pride, and the feeling of pride can obscure the love of the good. In pity, we compare ourselves to someone or something that is worse off. When Emile learns pity, he is described as putting himself in the

position of the suffering animal (**EE**, 222-23). This self-separation is characteristic human nature. Pity is *sweet* because, like with pride, we win the comparison. It's better to pity someone than to envy them. In envy, we put ourselves in the place of the other and find that they are better off. Envy turns the experience of something good into something *bitter*. The dictates of conscience become corrupted by the interplay of self-awareness and comparative judgment. This is how Rousseau can say that the conscience is both innate and born of the relation with one's fellows. And, in virtue of this, he can paradoxically maintain that everything god creates is perfect, even though some of God's creatures are inherently corrupt. Perfectly made, we transform ourselves through an act of self-separation or alienation. We can now see that Rousseau's analysis of pride and alienation is rooted in theology and not empirical anthropological claims about the superiority of indigenous ways of life.

So how are we supposed to get back from this fundamental corruption? It follows from Rousseau's view that there are two somewhat reliable guides to telling the good from the bad. These are *conscience* and *nature*. Fundamentally, the conscience feels the good. Rousseau uses a variety of metaphors for our access to the good. In the space of two pages in an essay on Polish political reforms he describes conscience as a *voice* speaking to the heart, a *sense*, and a kind of *digestion*. He mixes the latter two metaphors in an important passage: "God forbid I should think it necessary to prove here what everyone with a modicum of good sense and stomach senses!" (**CGP**, 6.5). The law of nature is better felt than expressed. He attributes religious disagreement to the attempt to "make god speak" (**EE**, 295).[136] To whatever extent Rousseau's

---

[136] "As soon as peoples took it into their heads to make God speak, each made Him speak in its own way and made Him say what it wanted. If one what listened only to what God says to the heart of man, there would never have been more than one religion on earth."

law of the heart may be referred to as a principle or a divine voice, it is not expressible in any human language. It seems clear that he has in mind some kind of feeling. And this makes sense given the corrupting role of will and intelligence in Rousseau's theory of mind. Nature (excluding human beings) can always be expected to be good, since it was created perfectly by God. However, we cannot always be expected to recognize the goodness in nature. If one has a good stomach, they will have a taste for it, but if your stomach grows sour from pride alienation, you may find yourself fixated on finding fault with creation and distracted from the beauty and original goodness of nature.

Since we can be confident that we *would be* good if we could be entirely natural, Rousseau proposes what he calls "negative education." "Prevent vice from arising and you will have done enough for virtue" (**CGP**, 191). Of course, this is far too simplistic. Concerning a child's physical growth, Rousseau is content to allow nature to take its course. His complaints about swaddling and coddling seem to reflect a view that that nature always knows best. The education of the body is negative insofar as children are allowed to grow naturally, but socialization is more complex. Preventing social corruption requires a much more active approach. But this activity should always be understood as protective and preventative. Emile's tutor (Jean-Jacques) doesn't shape Emile's character. He creates the conditions under which it can take shape on its own. To the greatest extent possible socialization should proceed according to the natural progress of emotions. Distinctively human emotions like pride, pity, and the fear of death shouldn't be taught before their time. Emile is sixteen before understanding that others suffer in the way that he does. Prior to this, he has "neither felt nor lied." Of course, he has *felt,* otherwise he couldn't have suffered, but he hasn't felt in that unique way in which human beings

do, "by transporting ourselves outside ourselves" (Emile, 222-23). Before his first experience of

pity, his feelings are natural and unified. Rousseau insists on the importance of maternal

breastfeeding:

> The child ought to love his mother before knowing that he ought to. If the voice
> of blood is not strengthened by habit and care, it is extinguished in the first years
> and the heart dies, so to speak, before being born. Here we are from the first steps
> outside of nature (**EE**, 46-47)

The insistence on maternal nursing is not about nutrition. It is through interaction with family

that sociality develops naturally. The conclusion of the book makes it clear that Emile's

education is not actually ideal. The point of *Emile* is not that every child should have their own

tutor. When Emile informs an aging Jean-Jacques that he will soon be a father, he congratulates

him on a successful job as a tutor and recommends that he retire (**EE**, 480). The message here is

clear. It is now Emile's turn to assume the role of tutor and bring up his own son as Jean-Jacques

educated him.

The family is the foundation of the state. Rousseau calls the family "the small

fatherland," and through it, the heart attaches to the state. One becomes a good citizen by being a

good husband, son, and father (**EE**, 363). The political realist principle that obedience is

grounded in fear is false. The family naturally expands the scope of self-interest. The first step

"outside of nature" is the natural affinity that a baby has for its mother. At this point, self-interest

isn't even possible because the infant isn't self-aware. When self-awareness occurs, the natural

love of the good will lead to self-love, but if things are going as they should, the child will

understand him or herself as a member of a family before understanding him or herself as an

individual. Rousseau describes puberty as a kind of "second birth." We are born first for our

species and only later are we "born for our sex" (**EE**, 211). Sexual desire naturally leads to

reproduction, but it also leads to the creation of a new family. The natural love of one's family

makes pure self-interest impossible. Self-interest is always contaminated by familial love, which

places the interests of others – wives, husbands, children -- within the scope of self-interest.

Emile's son's interest will be his own self-interest. Thus, the family is the first and most natural

way in which the fundamental problem of *The Social Contract* is resolved. It is a form of

association that allows each to unite with all while obeying only oneself (**SC**, 1.5.6). Love binds

the first and most natural form of association. Through voluntary marriage, one is transformed

from a man or woman to a husband or a wife. With this transformation, conscience is recast in

terms of relations. This is supposed to affect the way in which things make you feel. What was

once an innocent indiscretion becomes a heartbreaking betrayal; acts of kindness become

expressions of familial devotion. The child finds itself in a role before even knowing it.

Becoming a son or daughter is not like the voluntary transition from bachelor to husband. The

son should love his mother before he even knows her. In effect, the original conscience is

disposed to this affinity, and the family works toward the completion of the work of nature by

guiding the child's first steps beyond nature. If all goes well, then the child will have an

appropriate appreciation for the interests of others and a certainty of self that forestalls the

corrupting influence of pride.

One has to be careful not to misunderstand this point about the family. Rousseau is no

paternalist, but his rejection of paternalism is subtle. The family *is* a kind of society. He calls it

the oldest kind and says it's the only natural one (**SC**, 1.2.1). But in the *Discourse on Political

Economy,* he tells us that the family and the state have almost nothing in common, except that the

leaders of each have an obligation to care for the happiness of the members (**DPE**, 7). This latter

point, however, must be tempered with an observation about patriotism. Rousseau is clear that

citizenship involves a change of heart. The city, like the mind, has an active and a passive aspect.

As *sovereign*, it is active. As *state*, it is passive. The people are *citizens*, insofar as they

participate in the activity of the *sovereign* and *subjects*, insofar as they are passively affected by

the laws of the *state* (**SC**, 1.6.10). In this way, the double-relation of the state to itself parallels

the double-relation within the individual. To become a citizen, one must give oneself over

entirely to the body politic. The transformation of one's own self-interest into the interest of the

social body parallels the transformation from individual to family-member. The scope of self-

interest is expanded, and the conscience is redefined in terms of the relation to others. In this

way, love of country (patriotism) plays the role in the state that natural love plays in the family.

Therefore, we must not understand the rejection of paternalism as glib assertion that the family

and the state "have nothing in common." Paternal authority in the family is grounded "in the

nature of the matter." That is, in what Rousseau takes to be the significant natural differences

between men and woman and the necessary dependence of children on their parents (**DPE**, 4).[137]

So, what differentiates patriotism from familial love involves the character of the institutes that

define the relations among the members. The state is much bigger than the family. If nothing

else, it's simply impractical for a paternalist government to be so closely involved in the affairs

of each individual as a parent is involved in the life of a child. The children in a family naturally

depend on the parents. Citizens are equals because the dependence they have on one another is

social and therefore the subject to the legitimate limitations of the social contract. These

---

[137] I hope it goes without saying that this exegesis does not imply an endorsement of Rousseau's sexism. See note 11.

differences make familial love and patriotism very different, but they do not undermine Rousseau's basic insight: the sense of belonging to a family naturally expands self-interest to enliven *common* interest, which is a precondition for a just society.

Rousseau's egalitarian perfectionism is illustrated nicely by the character of the Savoyard Vicar. In *Emile,* the semi-fictional Vicar was to Jean-Jacques as Socrates was to Plato. Jean-Jacques describes his encounter with an inspiring role model who helped turn him toward the good, just as his youthful idealism was giving way to bitterness and resentment. The Vicar shows up as a kind of genius whose guidance liberates Jean-Jacques from the bondage of common opinion. So far, the Vicar sounds like a Socrates, leaving the metaphorical cave and returning with a higher wisdom that he then seeks to share with those still in bondage. With Socrates, the lowly masses are unwilling to hear this higher message, and they kill the wise man, all the while mocking what they don't understand. Socrates is represented as a kind of tragic hero with a sage-like wisdom and heroic self-mastery. Rousseau's Socratic hero, on the other hand, got booted from the priesthood because he couldn't keep his chastity vows. The Savoyard Vicar is the very opposite of an elite genius who raising himself above the masses. We are told that "his respect for the bed of others" made it hard for him to control his own desires (**EE**, 267). Had he been contemptuous of the laity, he might have been able suppress his own sexual urges—winning pride in exchange for gratification. Had he kept his chastity vows, society may have admired him, but he would have been no hero to Rousseau. In suppressing his natural urges, the priest gets things exactly wrong. There is nothing wrong with sexual desire, in itself. As feeling, it belongs to God's perfect creation. Socially, it plays an essential role in the formation of a family and, hence, the expansion of self-interest described above. If there is evil in sex, it is evil that we

add to it. The understanding risks turning sex into the servant of pride. This is why Emile's

negative education involves not allowing his sexual desires to get out of control before he's able

to handle them in a healthy way.[138] But the priestly vow of chastity is also a perversion. The

priest proudly compares himself to others and finds himself superior, not realizing that in

rejecting the sweetest of natural sentiments he is turning away from God and distorting his

relation to the rest of humanity.

We have seen that Rousseau sees the human condition as a kind of self-separation. We

have also seen that the highest good for human beings is a kind of self-understanding, whereby

one returns to one's original goodness in a knowing way. From Rousseau's starting point, we

have two guides as we attempt to know the good. First, we know that nature is good because it

was created by god. If we could follow the way of nature, we would do well. The tragedy of our

nature is that we can't simply follow that path. Second, sentiment is a reliable guide to the good

because as it is passive. The original conscience in its purity is a revelation directly from god.

But the 'word' of god is the corrupted the instant that we attempt to translate it to the human

understanding. This is a complicated picture. If you emphasize a different aspect, you get a

different interpretation of Rousseau. There is the *naturalistic* interpretation that emphasizes the

goodness of the natural. At its worst, this interpretation accuses Rousseau of advocating a

primitivism.[139] There is the *proto-Kantian* interpretation that emphasizes the activity of the will

---

[138] Volume One of Foucault's *History of Sexuality* describes the 19th century hysteria over childhood and adolescent sexual development. It's not hard to see how Rousseau's theory of education could be incorporated in such a crusade. But it's also not obvious that the concerns underlying this hysteria are misguided. Sex education certainly can be traumatic. As we will see, there is a fruitful way of objecting to the specifics of Rousseau's theory of education that proceeds on these grounds: if the result of education is an unhealthy obsession with (or fear of) sex, then it's likely that something is wrong with the educational approach and that it must be revised.

[139] Voltaire, August 30, 1755.

and understanding as distinguished from the passivity of the passions.[140] And there is the *romantic* or *mystical* interpretation that emphasizes the incorruptible and ineffable character of genuine feeling.[141] Each of these gets something right but suffers to the extent that it neglects the other aspects. If I'm right, the correct interpretation is the one that can bring these aspects of Rousseau's thought together by emphasizing the transformation in human nature and the achievement of a distinctly human perfection. Human nature is such that one begins life naturally and originally perfect, but the active faculties enable it to twist free of its original unity falling into a kind of self-deception or alienation. The self is divided against itself. The solution to this problematic is to follow the way of nature, which leads through one's engagement with others back to this original unity. Although this sounds fuzzy and mystical, what's really going on is quite ordinary. Through the love of one's family and country conscience is transformed. Self-love is expanded to promote the interests of others. This makes it possible to understand ones worthiness in terms of one's social role. You can be good husband, sister, citizen, soldier etc. If it goes wrong, you can end up like the bourgeois or the priest, hopelessly divided against yourself. If it goes well, you become what you've always been; you return to your original consciousness of the good with a deeper understanding. The father's natural affection for his children makes him love them whether he thinks of himself as a father or not. But through his transformative experience of becoming whole, wherein love is enriched by understanding. So, even though the object of the original conscience remains ineffable and unintelligible, the object of the conscience born of our relations with others can be approached in a knowing way. Through this

---

[140] Cassirer, 1945, 1989.

[141] Babbitt, 1919.

experience, we can enjoy the good (almost) as God does without becoming Gods ourselves. Alienation is a necessary precondition for this enjoyment, and the goal of education is the resolution of this internal conflict. An education that facilitates consciousness's knowing return to its original unity will make possible humanity's highest perfection.

3.4.   Perfectionism and Sensitivity to Moral Truth

In Rousseau's account, virtue, happiness, and moral knowledge all cut in the same direction. It is our destiny to know the good. So, if we live rightly, we will be happier and better at discerning value. From this it follows that one's way of life is a legitimate consideration when determining the reliability of one's moral sentiments. A way of life that keeps one's attention close to the moral law written on one's own heart will make one's moral instincts more reliable. A way of life that clutters and obscures this deep wisdom will make one's moral sentiments unreliable.

Evolutionary debunking arguments are sometimes thought to undercut moral knowledge by showing that the genealogy of our moral beliefs is insensitive to objective moral truth. But we have to be careful about exactly what this means. Science fictional thought experiments about belief forming pills are misleading if we allow that the mind has the power to actively assent to, or even construct, moral norms.[142] First of all, nature doesn't make us assent to even our strongest sentiments. An alienated conscience may feel tortured by the poorness of fit between desire and virtue. Second, the native impulses that such thought experiments are supposed to replicate provide only part of the genealogy of moral belief. The bulk of the environment in which we operate is or once was created by human beings. Our innate impulses may be

---

[142] For a further discussion of the Napoleon pill example, see 2.2 and 2.4.

redirected toward purposes that entirely differ from those for which they originally evolved.

Finally, empirical research can only warrant the positive claim that our moral psychology is

sensitive to non-moral facts. It can't prove the negative claim that our moral psychology is

entirely insensitive to moral truth. If there is indeed a deep moral sentiment concealed by the

clutter of other non-moral, and maybe even anti-moral sentiments, then by shaping our

environment we may succeed in turning our attention toward the former. Thus we can expect

some cultural practices to help facilitate virtue and others to debilitate it.

In a recent essay on Darwinian debunking, Tomas Bogardus calls refers to what he calls

the "divine revelation" view of moral psychology. On this sort of view, our moral psychology is

sensitive to moral truths.

> And that is because: if the moral facts had been different, the moral testimony of a
> loving God would have been different. And if God's moral testimony had been
> different, on this view, then our moral beliefs informed by that testimony would
> have been correspondingly different.[143]

Rousseau's view is sensitive to moral truth for a similar reason, but we have to be very careful in

talking about the supposed 'testimony' of God. Rousseau's account is distinctive for at least two

reasons. First, the voice of conscience cannot be expressed verbally without being corrupted.

Thus, it resists being conflated with any principle. Truth preserving principles will be empty

tautologies, like, it's good to do good. What's good one is left to puzzle out through an honest

reckoning with one's own conscience.

Second, as an outgrowth of our freedom, the moral sense is anchored in the human

condition. Given freedom's centrality to moral thought, this gives us a principled reason for

privileging the verdict of sentiment when it accords with freedom. It would be hard to be in

---

[143] Bogardus, 2016: 643.

harmony with oneself if one's moral worldview undercuts the very basis of morality. Our love of freedom is an (imperfect) outgrowth of our love of the good. For these two reasons, Bogardus's summary of the divine revelation view doesn't apply here:

> So, one way to rationally maintain moral realism in the face of evolutionary debunking arguments is to accept Divine Revelation. Some skepticism will still be called for though: the facts of evolution serve Divine Revelationists a defeater for those moral beliefs that they take to have been based on the deliverances of their evolved moral faculty (e.g. incest is wrong, rape is wrong, etc.). But Divine Revelationists can in good conscience—and in full knowledge of the facts of evolution—maintain their moral beliefs they take to have been formed via divine revelation (e.g. incest is wrong, rape is wrong, etc.).[144]

That way of understanding the divine revelation view leaves no room at all for us to learn anything about our moral worldview from evolutionary inquiry, and I have already said that I think there is much to learn from the selective debunking of various moral instincts. Proponents of the wisdom of repugnance view—which treats disgust as a divine source of moral knowledge—can simply plug up their ears and ignore what science has to say about the origins of that emotion. The views that I'm defending here are more nuanced. These views hold that normative inquiry is fundamentally different from descriptive science. This creates space for distinctively philosophical inquiries concerning the nature of the good while allowing that the scientific investigation of human nature provides valuable information concerning the moral significance of certain emotions.

What the egalitarian perfectionist view must insist on is that there are (perhaps very rare) instances in which moral emotion rises to the level of a genuine encounter with the good. It belongs to our nature as free beings that we can be both alienated from ourselves and capable of

---

[144] Bogardus, 2016: 659.

profound moral perception. Our encounter with the good is both a fundamental feature of the human condition *and* heavily influenced by context. Custom can help or hurt us in this regard. Rousseau was a pluralist about the particulars of religion. Judaism, Christianity, and Islam all reflect imperfect efforts to translate "the book of nature," which is immediately available to conscience (**EE***,* 303-7). The dogmatic commitment to religion errs insofar as it conflicts with morality and what Rousseau calls "civil religion" (**SC,** 4.8).

To recap, I have mentioned three criteria according to which the verdict of the moral conscience be tested. The centrality of alienation and the human condition suggests privileging feelings that accord with human *freedom*. And the perfectionist standpoint implies that a good conscience will belong to the person who lives a good life. As such the verdict of conscience should be *reliable* and lead to genuine *happiness.*

Of course, when I put it that way, it sounds hopelessly naïve. Perhaps we can imagine a moral utopia in which virtue truly is its own reward and nobody cheats. And perhaps the inhabitants of this utopia can see how great it is with their perfect moral intuition. But what does it mean for people in the real world? One troubling implication of the perfectionist view is that the corrupt are unlikely to know that they are corrupt, since their consciences are rendered unreliable by corruption. The only way out of this conundrum is to start trying things out. Considerations of freedom, well-being, and reliability constitute legitimate *external* standards according to which we can test the unity of an account of the good life against the *internal verdict* of conscience.

Bogardus points to the examples of rape and incest in his discussion of the divine revelationist view. It seems that advocates of a divine revelationist view can just re-name any gut

response 'revelation,' but perfectionists must take the genealogy of the response into account. The inner verdict of conscience must fit with what we know about its origin as part of a broader external account entailed by the theory. Jonathan Haidt famously defends the thesis that moral deliberation is a kind of post hoc rationalization. In an essay called "The Emotional Dog and its Rational Tail" he argues that the problem with moral rationalism is that moral judgment is primarily caused by intuitions that we then justify, only after we've already made up our minds. To show this, he surveys respondents using questions designed to trigger emotional responses to moral issues. Bogardus's mention of incest above is probably intended to reference to fictional description of a case of consensual incest between adult siblings that Haidt used in one of these surveys.[145] The case is tailored so that standard moral objections won't apply. The sex is entirely consensual. Precautions are taken to prevent pregnancy. Neither participant is harmed emotionally or physically. Still, when respondents hear the description of the case they insist that it was wrong and struggle to find some, any, reasons in defense of their verdict. Since the standard objections are ruled out, these rationalizations inevitably fail. "Eventually, many people say something like, 'I don't know, I can't explain it, I just know it's wrong.'"[146]

Haidt's purpose in conducting this survey is to show that reasons play less of a role in moral inquiry than we tend to think. So far, this is entirely consistent with Rousseau's sentimentalist view. But it's tempting to go further and wonder whether these cases of moral dumbfounding suggest that the moral belief in question is unjustified. If respondents can't

---

[145] It's worth reiterating that while Haidt's examples have become fodder for some fascinating ethical debates, Haidt's intention in the essay was not to show that the moral beliefs reflected in everyday responses are false. His point is a descriptive psychological one. When the triggered, people respond emotionally first, and rationalize their emotional responses after the fact.

[146] Haidt, 2001.

provide good reasons to support their moral judgment, how can they be justified in holding that judgment? But it would be a mistake to draw this conclusion. The subject's gut response might still be sensitive to—or a reliable indicator of—moral truth, even if it has a non-moral origin and delivers the wrong verdict in certain cases. Just as Renaissance painting used perspective to create the illusion of depth, a well-tailored thought experiment can create the illusion of a moral fact. If the gut response gives a true verdict in most real cases, the reliabilist can end it there and ignore the question of whether the response is sensitive to moral truth. Assuming that most real-world cases of incest will not match up with Haidt's carefully detailed example, the example fails to show that the gut response is unreliable.

But perfectionism allows for a more robust response. It enables us to approach the question more holistically, considering it in its broader genealogical context. Part of this genealogical context includes facts that may be incorporated in selective evolutionary debunking argument. There is support for the hypothesis that our sentiments concerning incest evolved to promote greater genetic diversity. Reproduction between close genetic relatives could lead to higher rates of deleterious recessive traits and greater susceptibility to certain kinds of infectious disease.[147] A moral argument on these grounds would be suspect, morality is not generally thought to correspond with the genetic strength of the human species—a frightful prospect given not too distant historical examples, like eugenics, genocide, and miscegenation laws. And even if someone did accept a principle like this, Haidt's example stipulates that the act will not result in pregnancy. So, it seems that the evolved aversion is likely to be an unreliable guide to moral

---

[147] Lieberman et. al., 2003. Felser and Navarette, 2004.

value in cases like these. Perhaps, the subject has a negative feeling about the act, not because

incest is actually wrong but because of an evolved aversion to fitness diminishing behavior.

I'm willing to grant that incest aversion is evolved and natural, but I reject the inference

that moral feeling is, therefore, insensitive to the moral truth of the matter. I don't think it's far-

fetched to think that a major part of the reason we find incest repulsive admits of this sort

naturalistic evolutionary explanation. To know this is to be aware of a natural bias against that

sort of behavior. But this bias is not a complete genealogy of any individual's felt response. The

perfectionist view puts us in the position to consider the issue more holistically, and when we do

this, we find that the innate aversion is only part of why we feel the way we typically do about

the issue. The connection that the incest aversion seems to have to the disgust response should

make us suspicious of the urge to cleanse society by casting out some supposed impurity. We

might go one step further and say that the psychological biases that stem from the logic of

contamination makes it especially important that we be attentive to the implications of the stigma

surrounding issues like sexual assault and HIV.

But a complete genealogy of an individuals response to one of Haidt's examples will

include cultural components as well. And this part of the story may be sensitive to a deep moral

truth. If Rousseau's description of the role of the family in moral development is granted, it

becomes apparent that incest is damaging to a certain kind of valuable interpersonal relationship.

We have seen that on Rousseau's perfectionist view family is crucial. Familial relationships are

unique. They channel natural sentiment and pave the way for other social bonds. If persons have

unconditional value, it's unrealistic to think that we can appreciate every individual to the extent

that is warranted by their nature. Countless people die every day, and if we got upset over each of

them, there'd be no time for anything else. Losing a close family member can be a personal

tragedy. In this way, familial relations facilitate an appreciation of the humanity in others. It takes

a long time to truly get to know someone and appreciate the depth and complexity of their

experience. My concern for the guy down the street is limited, but when I am reminded that he

has parents and siblings who care for him as I care for mine, it's easier to approximate his true

worth. Family can be a reference point for the unconditional value of humanity. It can also

facilitate self-understanding. Emile's son will know himself more deeply for the fact that he

understands himself as his father and mother's son. The good is mysterious; the role of a good

son is more transparent. The simple deeds associated with this role will get him acquainted with

the good without the impossible task of comprehending it. If things go well, he will grow into a

confident and autonomous individual. Self-assurance will make him less susceptible to the

corruptions of vanity and a better judge of what truly matters. The family relationships that make

this possible are decidedly different from romantic relationships built on sexual attraction. One

needn't be a puritan to allow that the desire for sexual gratification incorporates an element of

instrumentality that may distract from the unconditional worth of human dignity. So, we might

conjecture that through the social institution of the family, natural sentiment—including the

incest aversion—is redirected toward a network of unique relationships that promote moral and

emotional maturity. From this perspective, incest is seen to be wrong for the damage it does to

these relationships.

    Of course, the victims of abuse should not be slandered as irredeemably corrupt or

damaged, but it seems reasonable to think that something very valuable has been taken from

them. What makes familial relationships valuable is difficult to express, but respondents to

Haidt's survey's may express the point well-enough by the mere invocation of the relationship: "that's *his sister*!" If this invocation isn't sufficient, a more detailed story about the role of family in moral development can be offered on perfectionist lines. Notice that this account has at least as much to do with constructed social roles as it does with natural instinct. An analogous plea could be made in case involving a trusted mentor. It might be stronger or weaker depending on how crucial the relationship is in fostering personal growth. We shouldn't be surprised that respondents engage in shoddy ad hoc moralizing when asked to explain what makes this sort of thing wrong. If all goes well, there's little practical reason for anyone to know *why* it's wrong.

And when Haidt specifies that neither sibling was emotionally hurt by the encounter, the most obvious reason is excluded *ex hypothesi*. There are no grounds for demanding that moral feeling get this sort of case right. And even though the example is *logically* coherent, respondents (especially those with siblings) are likely to notice a poorness of fit. People are rightly surprised that siblings could do this without feeling profoundly conflicted. Knowing that the two are brother and sister gives the respondent good evidence that this just isn't how their relationship works. And when it's described as though the sex was no big deal, the sense that something has gone wrong with their relationship is compounded. The potential consequences—not realized in this hypothetical case—are disturbing too. Incest is associated with other dysfunctional family relationships. Families in which it occurs tend to exhibit increased levels of social isolation and relationship distress, and when the children grow up, they tend to exhibit increased levels of sexual discord in their adult relationships. Sibling incest often involves the abuse of a perceived authority by older over younger siblings and commonly precedes other patterns of abuse.[148] So,

---

[148] DiGiorgio-Miller, 1998. Saunders et. al., 1995. Tidefors, 2010. Worling, 1995.

while the feeling is hard to separate from the evolved aversion, it seems entirely plausible that the "it's your sister/brother!" response is the function of a well-calibrated and reliable moral conscience. What's wrong with the act is made more intelligible by focusing on the role of family in socialization and moral development.

By considering the internal verdict of conscience in perfectionist terms, we bring sociality and education into the genealogy of moral feeling. In so doing, we are better able to see how part of our spontaneous reaction to Haidt's thought experiment gets something right. This doesn't vindicate disgust as an innate moral emotion, but it puts this moral dumbfounding in perspective. On the perfectionist model, we can assess the reliability of the inner verdict of conscience more holistically, in terms of external consideration which may include a genuine sensitivity to moral facts. Egalitarian perfectionism puts freedom and alienation at the center of moral life and prizes relationships promote a healthy engagement with this feature of the human experience. Looking at things this way fosters appreciation for traditions and institutions like the family, but it doesn't justify closed-mindedness. One can appreciate the value of a particular institution without supposing that there is one eternal way of life capable of tending to humanity's deepest spiritual needs. Ways of life are historical; they respond to material conditions. The genealogy of morals is sensitive to moral truth, but not *only* moral truth.

This is how we should look at the extremes to which Rousseau takes his naturalism. If perfectionism is true, we will often err in our moral judgments. Rousseau frequently extols the virtues of a life in harmony with nature, expressing contempt for virtually all forms of progress. Moving back to the farm may seem like a nice alternative to vain titles and status seeking, but it can't be the one and only model of the good life. When a quiet life allows one the time and

clarity to reflect on those relationships that make life worth living, the simple life seems to have a lot going for it. But when Rousseau makes the perverse assertion that the victims of the Lisbon earthquakes brought it upon themselves by living in tall buildings, his callousness is absurdly discordant with genuine moral sentiment (**OC,** IV, 1061). Rousseau may have thought this extreme worldview would lead humanity to genuine happiness and freedom—a radical utopian liberationism. I don't doubt that he meant it when he said things like this. And there is no *a priori* proof that such a view is wrong. But *he was wrong*. And we can see this when we consider the verdict of his conscience in light of its genealogy and broader perfectionist worldview. It is implausible to think that such extremism will promote human flourishing and moral virtue. Callous moralizing directed toward the victims of tragedy reflects an unreliable moral conscience lacking in compassion. The arbitrariness of justice by natural disaster reflects an insufficient appreciation of human freedom and equality. In the world of the perfectionist ideal, Rousseau wouldn't have felt such contempt for earnest human endeavors. But, in the world in which he actually lived, he felt a profound a need to attack the pretensions of his age with every weapon available. In the heat of his exchange with Voltaire, his crude Tower of Babel metaphor got away from him. That's a shame because one of the chief virtues of his theory of alienation and education is that makes the inner verdict of conscience less mysterious. Value judgments don't need to fall back on crude biblical references.

3.5.   Conclusion

In this chapter I have argued against the primitivist interpretation of Rousseau's moral philosophy in favor of what I call 'egalitarian perfectionism.' The view rests on an account of the

genealogy of morality that incorporates social elements but needn't deny nativism about moral emotions. Furthermore, it facilitates a helpful way of assessing the immediate verdict of the moral conscience on perfectionist grounds. The reliability of the inner verdict of conscience depends on external factors on how well it coheres with the genealogical account. The details of Rousseau's own view require substantial revisions, especially where he leans on an extreme version of moral naturalism. I have no interest in defending his sexism or his comments about doctors. But the egalitarian perfectionist view provides compelling resources of assessing evaluative perception.

4.   CONSCIENCE AS THE AWARENESS OF OBLIGATION

4.1.   Introduction

Rousseau's egalitarian perfectionism provides a holistic perspective from which evaluate the inner verdict of conscience. From that perspective we can vindicate common sense objectivism without rejecting the scientific approach to the investigation of human nature. Still, a serious problem remains: Rousseau's view has no principled way to distinguish between the genuine conscience from the passions. In section 4.2, I will outline this problem by bringing Rousseau's view into conversation with the contemporary sentimentalism of Jesse Prinz. Then I will attempt to explain the basic problem in terms of what philosophers sometimes call—following Wilfred Sellars—the myth of the given.

By way of response to this problem, I will argue (in 4.3) that Kant's robustly cognitive account of perception provides the basis for a principled distinction between conscience and the mere passions. By placing autonomy at the center of moral philosophy and considering the logico-phenomenal structure of the awareness of unconditional obligation, Kant is able to show that the feeling of conscience is characterized by a paradoxical sense of awe. This sense of awe is paradoxical because—as sense that one's consciousness of obligation makes one the ground of value—it is both elevating and humiliating. The moral agent is at once *honored* at being called to an unconditional project of self-perfection and *humbled* by the absolute inescapability of failure. It is this encounter with 'the good'—not as an external object, but as the inner ground of all possible obligation—that raises humanity infinitely above the rest of nature.

I conclude (in 4.4) the chapter by drawing on the *Lectures on Pedagogy* in order to differentiate Kant's account of awe from one found in contemporary empirical psychology. Dacher Keltner and Jonathan Haidt interpret awe as an emotion grounded in our hierarchical evolutionary history. This gives awe the character of instinctual subservience. In Kantian terms, this awe is pathological. Kantian awe, on the other hand, is practical. It involves the rational disposition to recognize the absolute and equal worth of all human beings. If there is a natural instinct for servility, Kantian awe overcomes it. This can be seen from Kantian remarks on moral education, remarks which derive additional support from Serena Chen's work on the pedagogical benefits of self-compassion.

4.2.   Why *this* Feeling? Sentimentalism and the Myth of the Given

Kant was mortified by the "deranged" prophets of fanaticism, with their "presumed immediate inspiration from the workings [*Mächten*] of the heavens" (**EMH**, 2:267). In addition, Kant's critical turn brought with it a rejection of Francis Hutcheson's moral sense theory and any kind of empirical approach to moral philosophy.[149] So, one might think that Rousseau's sentimentalist theory of conscience—as a kind of a divine voice within the human heart—is exactly the sort of view Kantians should reject. But Kant's relation to Rousseau is notoriously complex.[150] In the same essay in which he calls fanaticism "deranged," he praises Rousseau for his "enthusiasm." With out enthusiasm, Kant tells us, "nothing great has ever been accomplished." And he holds Rousseau up as an ideal to which the doctors of the Sorbonne ought to strive (**EMH**, 2:267). Kant's take on Rousseau remains puzzling, and it would go beyond the scope of this project to

---

[149] Redding, 2007: 161-66.

[150] Cassirer, 1945: 1-18.

try and sort this out completely. But in considering the philosophical problems inherent in Rousseau's view and how Kant's theory of perception rises up to meet them, we may develop a certain appreciation for Kant's ambiguous admiration for Rousseau.

The worry that Rousseau's view may lead to fanaticism is well-founded. It is grounded in a problem with Rousseau's theory of perception that can be characterized as a version of the myth of the given. In the previous chapter we addressed the problem of fanaticism in passing when it was noted that there are profoundly immoral world-views that can be held without contradiction. The response there was that we needed to look at these ways of life *honestly* and ask whether we can *sincerely* accept their implications. So, we can ask whether, for example, plantation owners in the antebellum south sincerely believed that the abuses of slavery that followed from their explicitly racist worldview were morally permissible? It's likely that at least *some* of them were deluding themselves—an honest appraisal of the worldview finds numerous contradictions. Maybe a campaign of abolitionist rhetoric would have been enough to revolutionize their moral thinking.[151]

But this response is unsatisfying. We cannot rule out the very real possibility that even the most monstrous abuses might be accepted sincerely by a consistent moral agent. An easy way to conceive of this is to imagine an agent so deeply committed to an immoral practice that he or

---

[151] Stanley, 2016: 63-66. I will have more to say about moral revolutions in Chapter Five.

she is willing to adjust their entire system of moral and theoretical beliefs in order to justify it.[152]

We can call an agent willing to adjust all the elements of his or her system of belief in order to

support an immoral principle a *fanatic*. And from the point of view of the Rousseauian

conscience, it's hard to say definitively which feelings are apt to the good and which are not. It

seems as though this sort of deep-wisdom account of conscience can only promise that, deep-

down, we all have the ability to hear the call of the *genuine* moral conscience. Of course, this

may be exactly what the fanatic wants to say he or she is doing. In what follows, I will sketch

this problem by bringing Rousseau's view into conversation with the contemporary

sentimentalism of Jesse Prinz and explain the major weakness of this kind of view in terms of the

myth of the given.

4.2.1.   Sentimentalism

Rousseau's account of conscience can helpfully be compared to the contemporary

sentimentalism of Jesse Prinz. While Prinz doesn't appeal to Rousseau directly, the views have

quite a bit in common. They are both broadly sentimentalist, and they both insist on a radical

degree of the cultural variation in human nature. However, Prinz rejects what he calls

'externalist' accounts of the referent of moral judgment. If Prinz is right, it is hard to see how

---

[152] One might think here of pro-slavery interpretations of scripture, which seem so implausible to us today. Forbes (1998) gives this issue an admirably academic treatment: "To the contemporary reader, the Bible's position on slavery is ambiguous and contradictory. On the one hand, a host of passages paint the institution in a negative light and impose stiff sanctions on certain practices related to it. Most notably on the anti-slavery side of the ledger, the book of Exodus provides a classic paradigm of the passage from slavery to freedom. Deuteronomy 24:7 condemns 'man-stealing' as a capital offense, and another passage forbids the Israelites to return an escaped slave to his master (23:15-16). On the pro-slavery side, however, both the Hebrew Scriptures and the New Testament portray societies not only tolerant of slavery but economically dependent on slave labor; the institution is never condemned outright. In a pitched battle of exegesis, as generations of controversy would prove, pro- and anti-slavery controversialists could generally match each other verse for verse and interpretation for interpretation" (71). Quoted in Meacham, 2006: 325.

sentimentalism could be consistent with objective knowledge about moral facts, but I will show

that the challenge raised by Prinz does not undermine Rousseau's view.

Once this is clarified it will be easy to see that the real problem with Rousseau's view

actually has to do with his theory of perception. It is a version of the myth of the given.

Rousseau's broadly empiricist theory of perception—and, by extension, the perception of the

good—places the given object of sensation in the incoherent position of being both *a brute fact*

(devoid of any inferential content) and *a source of knowledge* about its object. This account

makes it impossible to tell an intelligible story about how experience gives us access to its

object. Rousseau doesn't run away from this problem. Instead, he is clear that the good cannot be

made intelligible, that we distort it when we put it into words.[153] But without an intelligible

account of the way in which conscience is able to access the good, Rousseau is unable to say that

this, rather than *that*, sensation springs from the authentic, uncorrupted moral conscience.

Rousseau's appeal to conscience posits a mysterious awareness of the good, and this

awareness becomes distorted the instant we attempt to make sense of its object. Jesse Prinz's

view is an interesting contrast case because Prinz is a sentimentalist about moral judgment who,

like Rousseau, places a heavy emphasis on the role of culture in shaping human nature. These

two facts should make Rousseau a natural ally. However, when it comes to the issue of moral

facts, Prinz insists on projectivism, arguing that moral facts are always response dependent.[154]

The problem with this combination of views is that it leads to a radical form of relativism that

conflicts with common sense objectivism. For, if the truth value of moral judgments depends on

---

[153] See section 3.2.

[154] Prinz, 2006: 34-35.

the actual human responses *and* if those emotional responses are substantially altered by cultural conditioning, then moral truth is straightforwardly culturally determined. In a culture that approved of exploitation, the claim "exploitation is good" would be rendered true by its context. The claim that a culture rejecting exploitation is morally superior to the former remains is untrue (perhaps, meaningless). Prinz's view threatens to vindicate a crass form of moral relativism. While similar, Rousseau's view avoids this unhappy consequence by insisting what I think few people would deny. Those who are culturally conditioned to see exploitation as an evil view the matter rightly, while those who disagree view the matter from distorted perspective. A culture that approved of exploitation would not just be different; it would be wrong.

Prinz doesn't attempt to satisfy the moralist's complaint. Instead, he argues that externalism is false. For Prinz, externalism is a thesis about the referent of moral judgments. If externalism is true, then there is some object—so to speak, 'out there'—to which true moral judgments correspond. Rousseau's view is an externalist on this description, since God speaks to us from the outside through the medium of sentiment. So, if Prinz's argument succeeds in refuting externalism, Rousseau's view is false. But I do not think that Prinz's argument succeeds against a view like Rousseau's. While Prinz provides powerful empirical support for his sentimentalist thesis that moral dispositions just are emotional dispositions, it is beyond the scope of empirical investigation to determine the ontology of the facts to which such dispositions correspond.

On Prinz's view, moral judgments are the expression of sentiment. Sentiments refer to response dependent properties, namely "the property of causing certain reactions in us."[155] So far

---

[155] Prinz, 2006: 34.

so good. Rousseau's account of pride is an example of this. The recognition of goodness associated with oneself and the approval thereof combined with the comparative judgment when considering others creates a feeling of superiority. If we talk about the referent of that feeling, it makes as much sense as anything to refer to its functional role in a causal chain beginning with a physiological response to stimuli and leading to a first-person phenomenal experience.

But Prinz goes further in insisting that we *project* this experience onto the world "in just the way we assume blue is a feature of surfaces."[156] Moral judgments, as emotional dispositions, do not refer to anything outside. He calls this feature of his view internalism. This is not to be confused with the expressivist thesis about moral language. Prinz's theory of moral language is cognitivist, at least to the extent that there is a fact of the matter concerning moral judgements. "If Coltrane never thrills me, I am being disingenuous when I claim to be a fan. Likewise, if I am never outraged by gender discrimination, I am paying lip service to equity."[157] The causal chain that leads to the emotional response is something real, and it can be studied objectively by studying the psychology of the individual. To this extent, there is a truth of the matter when it comes to moral claims. If the neurological effect of listening to Coltrane were the same as that of watching paint dry, it would be false to describe it as thrilling.

Furthermore, people really do make these sorts of false evaluative judgments, even about themselves. For example, behavioral research has shown that individual judgments about how good wine tastes are substantially affected by the taster's beliefs about how much the wine

---

[156] Prinz, 2006 35.

[157] Prinz, 2006: 38.

costs.[158] What's especially interesting about these studies is the fact that fMRI scans support the hypothesis that the taster's *actually are* experiencing greater pleasure when they taste the wine that they believe to be more expensive—even when tasting the exact same wine. If belief can affect sensation in such a way as to alter the perception of precisely the same object, it seems to follow that the conditions of perception are misleading. Under the right conditions, one should be able to perceive that the wines are identical or at least be able to rate them as comparable. Here we have a case in which the truth of the matter clearly corresponds to an external object, but the conditions under which the evaluate judgment is made corrupt the evaluative judgment. It's natural to think that the taster's pomposity or insecurity influenced their evaluation in a misleading way. A corrective might be to cultivate a virtue of epistemic courage or create conditions more amenable to objective deliberation. We do this in instances in which the stakes are high—like when we sequester jurors in order to insulate them from public bias. Empirical investigation supports the claim that moral judgments are tightly linked to moral emotion, but it cannot determine the extent to which such judgments answer to external facts. In some cases we really can be confident that there is an external evaluative truth. The exact same wine in a different bottle is just as good, and this is true regardless of what the taster *thinks* it costs. The taster is simply being mislead by the context. If you told them that the wines were the same, they'd be embarrassed. Prinz would have us interpret this as showing that it's all in the taster's head. But the empirical study of perception cannot establish this. It is just as consistent with the facts to say that there are more or less optimal conditions for judging evaluative truth.

---

[158] Plassmann et. al., 2007. Goldstein et al. 2008. See discussions in Henrich., 2016: 264-66 and Bloom, 2010: 45.

Prinz defends a broadly 'nurturist' account of human nature. On his view, emotions that appear natural are often best explained in terms of cultural context. He argues, for example, that observed differences in male and female responses to the idea of sexual infidelity are better explained by the dominance hypothesis than evolutionary psychological explanations that interpret male jealousy as an evolved response.[159] If this is so, it would seem to follow that there are better or worse social conditions under which to evaluate questions of gender discrimination. It seems quite plausible that internalized narratives of male dominance corrupt our everyday judgments about women and the relative value of their social contributions. To this extent, Prinz's view seems of a piece with prominent feminist critiques of patriarchy and male domination. He shows that many supposedly natural differences are better explained with reference to culture.[160] One might expect him to follow this up with the assertion that societies that impose structures of male domination on women when there is nothing in nature to justify such structures are wrong. Aren't unjustified structures of domination objectively worse than more equitable societies? And isn't this an evaluative proposition that goes beyond our mere *projection*?

Prinz doesn't go in this direction. He argues that there is no plausible account to be given of how moral judgments—construed in sentimentalist terms—might link up to an external moral standard.[161] He considers two prominent normative standards: maximizing utility and consistency in willing. If sentimental moral judgments refer to these normative standards, then they refer either descriptively or causally. That is, if moral judgments are really about, say,

---

[159] Prinz, 2012: 343-47.

[160] Prinz, 2012: 213-38, 330-63.

[161] Prinz, 2006: 39-41.

maximizing utility, then all true positive moral judgments will either *describe* or *be caused by*

what maximizes utility. If the descriptive account of reference holds, then "x is good" will be the

same as "x maximizes utility." Prinz rejects this by appealing to Moore's open question

argument. Despite its problems, Prinz says it shows that "there is no analytic tie between

ordinary moral concepts and the descriptive concepts that designate the properties indicated in

the moral theories of Kant and Mill."[162] Furthermore, the properties indicated by these theories

don't reliably *cause* our moral sentiments either. He shows this by pointing out that such

normative theories tend to promote impartiality in moral judgment, while our ordinary moral

sentiments are typically quite partial.[163]

What should the Rousseauian make of this dissociation of sentimentally construed moral

judgment with an external moral standard? The answer, I think, is very little, unless the

Rousseauian holds to the implausible premise that ordinary moral judgements are always fully

apt. It's important to keep in mind that any honest account of moral judgment (and this is

especially true from the sentimentalist perspective) will have to acknowledge that the actual

lived experience of moral judgment involves an admixture of moral, nonmoral, and perhaps even

immoral emotions. It seems to me that all the Rousseauian needs to insist on is that there is a

kernel of moral truth at the heart of ordinary moral sentiment. This is the feature of Rousseau's

---

[162] Prinz, 2006: 40.

[163] Prinz, 2006: 40.

view that seemed to have most profoundly affected Kant.[164] Ordinary moral sentiment does have

a kernel of truth to it, and the reason for this is that the fundamental ground of morality is the

voice of God speaking directly to the heart. Divine inspiration gets distorted by human emotions,

such as pride. But even pride still contains a spark of the divine goodness, insofar as it is the

recognition of goodness and the desire to be associated with it. That's why Rousseau needs two

separate words for pride: *self-love* that is fitting of even a perfect being (since a perfect being

will love the good) and *pride* that compels imperfect creatures to reduce their peers in order to

feel the false satisfaction of being better off by comparison. A similar thing can be said of pro-

social emotions, like pity. Rousseauian pity contains the same wickedness implicit in pride. The

sweetness of pity stems from the fact that we compare ourselves with the object of our pity and

find ourselves better off. For creatures like us, moral sentiment is corrupted by an admixture of

competing emotions, but a spark of goodness remains. It's unsurprising that evaluative

judgments people make about wine correlate with beliefs about how much it costs. This reflects

an awareness of the judgment of others (market price as a function of demand) and also an

awareness of social status that we attach to its consumption. When moral emotions intermingle

with ideas about personal identity, they can be distorted by feelings about who we are and how

---

[164] Schneewind, 1998: 487-92. Schneewind quotes a famous passage (**AA**, 44:8) from Kant's notebooks, concerning the influence of Rousseau: "I am myself a researcher by inclination. I feel the whole thirst for knowledge and the curious unrest to get further on, or also the satisfaction in every acquisition. There was a time when I believed that this alone could make the honor of humanity and I despised the rabble that knows nothing. *Rousseau* set me to rights. This dazzling superiority vanishes. I learn to honor man and I would find myself more useless than the common laborer if I did not believe that this observation would impart to all else a value to restore the rights of mankind."

we size up. This is also true of emotions like empathy, that seem to be focused on others.[165]

Therefore, the Rousseauian should have no problem with the fact that there is no direct causal or

descriptive link between the good as it truly is and moral judgments construed in terms of

evaluative sentiment. It is entirely natural to suppose that this is something that would need to be

worked out in thoughtful, and ever growing, moral contemplation.

4.2.2. The Myth of the Given

Rousseau's objectivist sentimentalism is not undermined by the empirical worries that motivate

Prinz's internalism. But to say that the true moral sense is mysterious and difficult to separate

from the admixture of ordinary moral experience threatens to make the normative ethicist's

project endless and impossible. The good lies as if on the other side of an abyss, and our access

to it is always already contaminated by our fallen nature. It should be noted, however, that this

embarrassing mystery at the heart of Rousseau's theory of the good is neither surprising nor new

to him. One may be humbled by how unscientific talk of the moral conscience quickly becomes,

but this seems be the same embarrassment expressed by Socrates in the *Republic* when he is

forced to characterize the good using the famous sun analogy.[166] Properly understood, mystery is

---

[165] Bloom, 2016 Bloom's book is largely a selective debunking argument against affective empathy. He posits rational deliberation as a corrective. Notice, however, that he doesn't abandon appeals to moral emotion all together. He substitutes what he calls "rational compassion." On his view, the problem with empathy is its narrow focus. "Empathy is a spotlight focusing on certain people in the here and now. This makes us care more about them, but it leaves us insensitive to the long-term consequences of our acts and blind as well to the suffering of those we do not or cannot empathize with" (9). This sort of critique of empathy is consistent with the idea that it, or a similar moral feeling, is the *ground* of morality (the initial or most authentic experience from which more refined concerns originate), provided that we make room for reason in its practical application.

[166] Plato, 1992: 506d. "By god, Socrates, Glaucon said, don't desert us with the end almost in sight. We'll be satisfied if you discuss the good as you discussed justice, moderation, and the rest.
       "That, my friend, I said, would satisfy me too, but I'm afraid that I won't be up to it and that I'll disgrace myself and look ridiculous by trying."

not fatal to a theory; nor is it license for dogmatism or extremism. But, to the extent that there is a mystery at the heart of Rousseau's theory, we should attempt to clarify it as much as possible.

So what does genuine moral sentiment actually look like? It is an unfortunate feature of Rousseau's account that he seems to make it impossible to provide a clear description of moral feeling. At best, this sort of view can only point us to our own hearts and tell us to feel it for ourselves. But what does one feel? The best candidates for the fundamental moral emotion admit of strong counter-examples. Pity is often condescending. This sort of complaint may set off the following dialectical progression. One person proposes a specific emotion as a source of moral knowledge. Then a critic shows all the ways in which that emotion can be misleading. Then the first proposes a more nuanced alternative that does a better job satisfying our intuitions. At this point, the cycle typically repeats with the critic pointing showing ways in which the revised notion is misleads or falls short. A recent book on empathy provides a good example. The author does excellent critical work, highlighting the various ways in which 'empathy' can be epistemically misleading. But then he proposes 'rational compassion' as an alternative, acknowledging that compassion and reason can both sometimes be misleading.[167] Rousseau's theory of conscience avoids this progression by severing the link between conscience and any particular emotion. This also allows him to evade evolutionary debunking explanations that might reduce our moral emotions to mere adaptive responses. Since there is no particular emotion to be found, no selective evolutionary debunking argument can be made. A general debunking argument would undermine the view, but we saw in the previous chapter that these arguments grow weaker the more general they become. So, while Rousseau would not have

---

[167] Bloom, 2016.

understood the problem in Darwinian terms, he seems to have understood the philosophical need to protect his theory of conscience by making it resistant to naturalistic explanation.

But this mystery comes at a cost. The conscience is starting to sound like an empty place-holder. If we can't specify a particular feeling (or set of feelings) as the source of moral knowledge, it's hard to make sense of how feeling can be a source of moral knowledge at all. In the previous chapter, I attempted to defend Rousseau's view by appealing to a strategy of self-justification. We begin by trying out a perspective and see how the view itself holds up according to its own purported standards. This facilitates a substantial amount of revision, but there are many unacceptable views that it cannot explicitly rule out. The fanatics view holds up particularly well. If you're willing to take a limited set of principles as properly basic and jettison all other earthly concerns in order to preserve consistency, you may do so while justifying horrendous moral wrongs. For example, if the divine command is always morally absolute and God commands Abraham to kill his son, then this command supersedes all earthly concerns. So much the worse for Isaac's right to life and personal well-being. It seems as though the same feature that allows Rousseau's theory of conscience to evade naturalistic and causal criticism, makes his view susceptible to fanaticism.

The fatal flaw in Rousseau's view lies in his empiricist conception of sensation. His view is guilty of what contemporary philosophers, following Sellars, call the *myth of the given*.[168] As a problem in the philosophy of perception, one comes up against the myth of the given when one posits a non-inferential entity in perception that serves as the basis for an inference in a claim to knowledge. According to the objection, the idea of pure givenness is contradictory because it

---

[168] Sellars, 1997. See discussion in Redding, 2007: 22-29.

requires a non-inferential brute fact to do what only an inferential elements, like concepts, can do. One target of the objection is Bertrand Russell's theory of knowledge by acquaintance, according to which sense data serves as a brute fact from which one is able to make judgments about the external world. So, when you walk around a table viewing it from various angles, light is hitting your retina creating color perceptions. None of these perceptions is the table itself, only sense impressions.[169] This sense data is then taken to be a brute fact—something given. From brute facts like these, you are able to infer the existence of a real table. The problem with this view is the contradictory nature of these brute facts. Brute facts are not the stuff of cognitive inference. They just exist. But if they are not inferential themselves, then it's hard to see how they can serve as the basis for any inference at all, and in particular, it is hard to see how they can function as the basis of an inference that there really is a table. Either the contents of our sensations already contain at least some inferential content or they cannot serve as the basis for our judgment. Setting Russell's epistemology aside, it is easy to see how this spells trouble for Rousseau's theory of conscience.

In fact, things are worse for a theory of conscience than for an epistemological theory. While the latter might attempt to respond to the objection by denying that any *inference* is being made in encountering a given, the theory of conscience clearly does involve an inference from a given impression to a fact (or set of facts) about what one ought to do. I don't know if the strategy of denying that any real inference is being made gets the epistemologist off the hook, but the strategy doesn't even seem to be available to a theory of conscience like Rousseau's.

---

[169] Russell, 1959. The comments about Russell are merely meant to illustrate the philosophical issues related to the myth of the given. Whether these issues threaten to undermine Russell's actual views is beyond the scope of this chapter. I only mean to show that *Rousseau's* theory of conscience depends on a notion of givenness that is helpfully illustrated in these terms. For a treatment of the issues as they pertain to 20th century analytic philosophy, see Redding, 2007.

Rousseau's view treats the good—as the object of authentic moral perception—both as a kind of a pure sensation and a "law written on the human heart." But it can't be both. If it's a *law* written on the heart, it has inferential content already built into it; if it's pure *sensation*, then there is no inference to be drawn. This explains the extraordinary lack of guidance and susceptibility to fanaticism that stems from Rousseau's sentimentalist worldview. If we follow his pure sentimentalism to its conclusion, we find ourselves dangled out over an abyss, forced to decide moment to moment whether the path we have chosen is true to the deepest sentiments of the heart. But in the complete absence of inferential content, there is no way to specify the conditions that sentiment lays out.

If we go the other way and lean heavily on the inferential content of moral sensation we end up with a much different view. It's worth noting that Rousseau's own view sometimes veers in this direction. We have seen that his account of distinctively human emotions—like pride and pity—have a cognitive structure. In addition, Rousseau's argument against slavery—which is essentially the same as Kant's—seems to point toward a more rule based moral understanding. In his rejection of slavery, provides a speculative history of the origins of servitude and considers what might make it legitimate. He concludes that there is no way for slavery to be legitimate, because a free being could not give away his or her freedom without also losing his or her authority to commit to servitude. So, he considers the idea of legitimately enslaving a free being is fundamentally inconsistent (**SC**, 1.1-1.2). It is from this sort of reasoning that Kant derives the duties one has to oneself as a moral being. Lying, avarice, and false humility are wrong because they contradict "inner freedom." This, according to Kant, "is tantamount to saying that they make it one's basic principle to have no basic principle and hence no character" (**MdS**, 6:420)."

Likewise, Kant's rationalistic approach to morality provides an appropriately forceful response to the fanatic (**KpV**, 5:86-87). Kant is typically understood to be a moral rationalist. If this means that he derives the principles of morality from reason without appealing to the sense experience, we have seen that there are good philosophical reason that this is project is doomed to failure— either as a moral realism without foundation or as an empty formalistic constructivism.[170] But by rethinking the moral sensation in more cognitively loaded inferential terms, it may be possible to recast the moral conscience in such a way that combines the advantages of rationalism and sentimentalism. So, it is to the Kantian account of the moral conscience that we now turn.

4.3.   Obligation and Humiliation: Kant's Cognitive Theory of Conscience

Kant explains his doctrine of conscience in the Doctrine of Virtue (**MdS**, 6:399-403). He defines conscience as "practical reason holding the human being's duty before him for his acquittal or condemnation in every case that comes under a law." He provides this definition along with moral feeling, love of human beings, and respect as "prior aesthetic concepts of the receptivity of mind for concepts of duty in general." These are the natural predispositions that must be present for the mind to be affected by duty. "To have these predispositions cannot be considered a duty; rather, every human being has them and it is by virtue of them that he can be put under obligation." Respect is defined here as a subjective feeling caused by the moral law within. There is no obligation to respect oneself; rather, the moral law within "unavoidably forces from him *respect* for his own being." Love of humanity [*Menschenliebe*] is distinguished from benevolence [*Wohlwollen*]. The former is a kind of feeling and cannot be a duty; the latter is the

---

[170] See section 1.4.

duty to promote the well-being of others. Like conscience and respect, love of humanity is a feeling that is presupposed as a prior condition of receptivity to morality. Moral feeling is the susceptibility to pleasure or displeasure "merely from being aware that our actions are consistent or contrary to the law of duty." In addition, "any consciousness of obligation depends on moral feeling to make us aware of the constraint present in the thought of duty." Again, there can be no duty have moral feeling because consciousness of moral obligation presupposes this feeling.

These receptive concepts are not the ground of morality; they are the preconditions of consciousness of morality. That Kant must insist on this is clear from his distinction between *pathological* and *practical* love. The former is a feeling that we have by inclination. It cannot be commanded. Benevolence is the duty that we have to promote the well-being of others whether we have any inclination to or not. In the *Groundwork* Kant's tone is derisive, associating pathological love with "melting sympathy." When the bible commands that we love our neighbors and even our enemies, Kant is convinced that what's commanded is benevolence. His view is consistent in both the *Groundwork* (1785) and the *Metaphysics of Morals* (1797) that duty must be necessitated, and "to be constrained to take pleasure in something is a contradiction" (**G**, 4:399; **MdS**, 6:401-2)

If conscience is on the side of spontaneous and subjective receptivity, and duty is on the side of the objective moral law, then it's reasonable to ask: what role, if any, can conscience play in Kantian moral philosophy? Yet, in the Doctrine of Virtue, Kant seems to identify moral feeling, conscience, love of humanity, and (self)-respect as *necessary preconditions* of morality. We can make sense of this if we look at the *Critique of Practical Reason.* Here Kant also

associates moral feeling with the consciousness of freedom and the moral law, characterizing it

specifically as a feeling of humiliation that, in turn, generates respect for the moral law within:

> Now the moral law, which alone is truly objective (namely objective in every respect), excludes altogether the influence of self-love on the supreme practical principle and infringes without end upon self-conceit, which prescribes as laws the subjective conditions of self-love. Now in what our own judgment infringes upon our own self-conceit humiliates. Hence the moral law unavoidably humiliates every human being when he compares with it the sensible propensity of his nature. If something represented *as a determining ground of our will* humiliates us in our self-consciousness, it awakens *respect* for itself insofar as it is positive and a determining ground. Therefore, the moral law is even subjectively a ground of respect (**KpV**, 5:74).

The consciousness of one's own freedom necessarily generates a feeling of humiliation because

we are aware that the moral law commands universally despite any and all inclination. It is for

this reason that Kant insists on the teachers of morality to employ "a dry and earnest

representation of duty" as examples in moral education (KpV, 5:157).

> When we can bring any flattering thought of merit into our action, then the incentive is already somewhat mixed with self-love and thus has some assistance from the side of sensibility. But to put everything below the holiness of duty alone and become aware that one *can* do it because our own reason recognizes this as its command and says that one *ought* to do it: this is, as it were, to raise oneself altogether above the sensible world… (**KpV**, 5:159).

This comes on the heels of a quotation from Juvenal that sounds remarkably similar to the

thought experiment that Kant uses to introduce the famous *fact of reason* (**KpV**, 5:29-31,

5:158-59). Rather than being asked to perjure your neighbor upon threat of the scaffold, Juvenal

asks you to imagine that the tyrant Phalaris threatened to cook you to death in a giant bull made

of brass if you refused to perjure yourself. While Juvenal calls it the "greatest of all inequities to

prefer life to honor," Kant would obviously have us substitute *the moral law* for honor. To

perjure oneself to preserve your life, and Kant's view as well as Juvenal's, is "to lose, for the

sake of living, all that makes life worth living."[171] That pure practical reason commands you to tell the truth contrary to all possible incentive is the purest expression of the consciousness of one's own freedom. This consciousness of freedom--the fact of reason that grounds respect for the moral law by way of the feeling of absolute humiliation—is Kantian conscience and moral feeling.

What's most striking about this kind of 'feeling' is that it is a logical outgrowth of what seems to be an entirely cognitive notion of obligation. On Kant's view, the causal ordering is essential. There can be no merely natural disposition from which moral feeling extends. If love of humanity, for example, stemmed from a kind of natural sentiment, it is of no moral worth. The causal order is reversed. "The saying 'you ought to love your neighbor as yourself; does not mean that you ought immediately (first) to love him and (afterwards) by means of this love do good to him. It means, rather, *do good* to your fellow human beings, and your beneficence will produce love of them in you" (**MdS**, 6:402). A love springing from natural sentiment, as if by instinct, would be of no moral worth and inappropriate for the kind of moral feeling that Kant is describing. Nevertheless, love of humanity can be cultivated by practicing that benevolence to which we are morally obligated. Notice that on an empirical conception of moral sentiment (like Rousseau's) this would have to be the end of the story. Rousseau's passive sentimentality originating in divinely created nature can furnish only immediate pathological love. This would render the appeal to the fact of reason no different from other pro-social sentiment. On Kant's view, the feelings of humility and respect are "unavoidably forced on us" by consciousness of the moral law. The passions are mechanical and subject to the empirical conditions of causality. But,

---

[171] Noted in the Cambridge edition of Kant's *Practical Philosophy*. Gregor (ed), 1997: 632,n14. Juvenal, 1991: 72-3.

unlike most empiricists, Kant limits the theoretical investigation of the causal ordering of the physical world to the phenomenal world of appearances. The notion of an intelligible world admits of the possibility of unconditioned causality, a concept that has significance for us in the form of the moral law (**KpV**, 5:56). So, by comprehending the idea of human freedom in terms of the obligations placed on us by the moral law—that is, by recognizing that I *must,* and likewise *can*, do the right thing regardless of *any actual* physical obstacle that may make right action difficult—I recognize myself as linked to something so immense (universal obligation) that it forces a sense of awe in and of itself.

This distinctively Kantian feeling of awe is articulated in the pre-critical period and linked to the notion of the sublime. The early Kant seems to have been fond of Hutcheson's moral sense theory, and Kant continues to say good things about Hutcheson despite explicitly rejecting moral sense theory in the *Groundwork* (**G**, 4:442-3).[172] However, the connection between universal moral principles and sublime feelings of awe extends well into the mature philosophy of the critical period. In his *Observations on the Feeling of the Beautiful and Sublime,* Kant tells us that "the sublime must always be large." He describes "a mountain whose snow-covered peaks arise above the clouds." The sublime touches us in a way that is sometimes accompanied by a feeling of dread ("a great depth" like the Grand Canyon), but it can also be noble (mountains and pyramids) or magnificent (St. Peter's Basilica). Even expansive lengths of time, as in depictions of eternity, are sublime (Observations, 2:208-10). If immense size is paradigmatic for the sublime, then it is fitting that universal principles—bound by no natural limitations—are also sublime since they extend out beyond all possible experience. According to

---

[172] See also, Redding (2007), 161-66.

Kant "subduing one's passion by means of universal principles is *sublime*" (**OBS**, 2:215). This is related to the sublimity of tragedy, as opposed to comedy. In comedy, "love is not so grave." If there's trouble, the characters know "how to wriggle out" of it. But in tragedy, "love is melancholic, tender, and full of esteem." The characters display "magnanimous sacrifice for the well-being of one another, bold resolve in the face of danger, and proven fidelity" (**OBS**, 2:212). Unlike the charming and kindly affections of a beautiful comedy, the faithfulness of a person who remains virtuous in the face of tragedy is sublime. Foreshadowing his mature moral philosophy, Kant concludes "true virtue can only be grafted upon principles, and it will become the more sublime and noble the more general they are. These principles are not speculative rules, but the consciousness of a feeling that lives in every human breast and that extends much further than to the special grounds of sympathy and complaisance. I believe that I can bring all this together if I say that it is the feeling of the beauty and dignity of human nature" (**OBS**, 2:217). We can think of the recognition of the universally binding character of the categorical imperative as forcing a feeling of awe upon us as something that breaks out beyond the ordinary causal order of things.

Tragedy moves us because of the mismatch between the way things are and the way that they ought to be. It breaks open the causal ordering of the world and forces us to think beyond it, in the same way that the vision of an immense mountain breaking through the clouds displaces everyday spatiality and overpowers us. The moral law in its boundless character does same sort of thing to our everyday understanding of ourselves. It forces us to imagine an ideal version of ourselves, one who never succumbs to temptation but always does what the moral law determines that they ought to. The Kantian conception of a holy will that requires no

necessitation is so far removed from actual human experience that the very idea opens up a gap

so great that it inspires a certain sense of awe (**G**, 4:414). Insofar as the idea of a holy will comes

merely from the idea of a good will that is *determined* by the moral law, as opposed to

necessitated by a kind of moral effort, it can be said that the feeling of awe forced upon is forced,

not by pathological feeling, but something that might better be described as a pure logic of

freedom. In this way, the Kantian moral conscience—as an awe stemming from the

consciousness of freedom and the moral law—is no ordinary emotion but rather a fact of reason

at the foundation of human morality. We are at once humbled by what we are not and elevated by

the idea of what our rational nature makes possible for us.

So, we have seen that Kant's thoroughly cognitive theory of conscience points to a

characteristic kind of moral feeling that is grounded in the consciousness of moral obligation. It

is not a pathological feeling. It is not some kind of gut instinct for morality. Instead, it is a kind

of recognition that stems directly from the idea of normativity itself and an awareness of what

our freedom makes possible—and also obligatory—for us. We stand in awe of the moral law, and

this generates a characteristic kind of feeling which we rightly take to be a sign of the nobility of

the moral law within. But the fact. that this feeling is rooted in the rational cognition of

normativity provides the grounds for a crucial distinction. This truly moral feeling is not to be

confused with pathological emotions. Hence, Kant's view provides what Rousseau's view

couldn't: a reason for identifying *this* feeling as *the* moral feeling. It is this that allows Kant to

say—unlike Mendelssohn who was forced to make an appeal to the external standard of common

sense—that *reason itself* furnishes its own standard (**WHD**, 8:133-46).

4.4.   Education and Fundamental Moral Feeling

Two important things follow from what's been said about the moral conscience so far. The first is

what's just been said: conscience is no mere feeling. It has as its ground the phenomenal logic of

freedom and obligation. Another consequence is that this feeling points toward the ideal of

universal human equality. Kant's own racism and sexism notwithstanding, human dignity is the

centerpiece of Kantian moral philosophy.[173] This is illustrated in the account of the vocation of

humanity, at the heart of his theory of education. In the remainder of this chapter, I will discuss

Kant's theory of education in order to draw a stark contrast between his account of awe and an

account from contemporary evolutionary psychology. Where Keltner and Haidt see awe as a

function of our hierarchical natural history, Kantian awe bares witness to the universal equality

and dignity of all human beings. We experience Kantian awe, not because we are amazed at the

power of other human beings, but because we are conscious of the immensity of the moral ideal.

This will be made evident by a brief consideration of Kant's account of moral education.

4.4.1.   Moral Education and the Vocation of Humanity

In the *Lectures on Pedagogy*, Kant is not shy about declaring his hopes for education. They

involve nothing less than the achievement of the destiny of humankind:

> Education [*Erziehung*] is an art, the practice of which must be perfected over the
> course of many generations. Each generation, provided with the knowledge of the
> preceding ones, is ever more able to bring about an education which develops all
> of the human being's natural predispositions proportionally and purposively, thus
> leading the whole human species towards its destiny [*Bestimmung*] (Pedagogy,
> 9:446).

---

[173] For a brief discussion of the literature on Kant's racism, see the introduction above.

The human being, he tells us, is endowed with natural predispositions that he calls germs [*Keime*] which lead them to reach their destiny. The presence of such natural dispositions, but unlike animals that use their abilities naturally, by instinct, human beings must be educated [*Erzhält*]. He goes back and forth a bit on animals, expressing some fascination that "birds to not sing by instinct, but actually learn" to do so. And he describes an experiment for proving that this is so. But unlike other animals, "the human can only become human through education" (Pedagogy, 9:443).[174] Through education, progressively improved over generations, human beings develop to the point of achieving the destiny to which we are naturally predisposed. He laments the conditions of education in his own age invoking the Aristotelian notion of "second nature."

> With the present education the human being does not fully reach the purpose of his existence. For how differently do people live! There can only be uniformity among them if they act according to the same principles, and these principles would have to become their second nature (Pedagogy, 9:445).

With all this talk of perfection, unity, second nature, and destiny, it seems hopeless to try and avoid interpreting Kant's theory of education as a utopian project to unify and perfect human society, the universal human project of establishing a kingdom of ends on earth.

But this construal of education as the progressive achievement of human destiny needs to be qualified in two ways. First, it must be remembered that the Kantian notion of an *idea* belongs to *reflective* judgment, which must be kept distinct from *pure concepts* of the understanding. The pure concepts of the understanding are the conditions for the possibility of all experience whatsoever, and they necessarily determine the relation of objects in space and time. On Kant's

---

[174] Ibid., 9:443. The experiment involves stealing eggs from a canary and substituting sparrow eggs, then isolating the little test subjects from other sparrows. The result, he tells us, is that the sparrows will learn to sing like the canaries. I have not looked into this.

view, there is no experience absent, for example, cause, and the principle of succession in time outlined in the second analogy holds universally for all experience without exception. The idea of reflective judgment is different. "An idea is nothing other than he concept of a perfection which is not yet to be found in experience." As an idea "it is by no means impossible," and "the idea of education which develops all the human being's natural predispositions is indeed truthful" (Pedagogy, 9:444-45). The idea of human nature can give us guidance concerning the nature of human perfection and, therefore, tell us something about what a better world would look like, but it does not permit us to make bold predictions about when and how human destiny will be achieved. A good example of this is Kant's essay *Toward Perpetual Peace*, he appeals to what he calls (elsewhere) 'unsocial sociability' to explain how nature "guarantees perpetual peace through the mechanism of human inclinations itself" (**AA**, 8:368).[175] Kant's view is essentially that war will become too expensive, and the "spirit of commerce" will prevail by appealing to human interest. It will prove too difficult for any individual despot to conquer the world and impose a "universal monarchy," and as nations develop, promoting peace will be in the interest of trade. But he qualifies this 'guarantee' as "admittedly not adequate for predicting its future (theoretically) but that is still enough for practical purposes and makes it a duty to work toward this (not merely chimerical) end" (**AA**, 8:364-68). Thus, we can think of Kant's

---

[175] Compare to Universal History, 8:20: "The means nature employs in order to bring about the development of all its predispositions is their *antagonism* in society, insofar as the latter is in the cause of their lawful order. Here I understand by 'antagonism' the *unsocial sociability* of human beings, i.e. their propensity to enter into society, which, however is combined with a thoroughgoing resistance that constantly threatens to break up society." Kant's history of warfare is both fascinating and odd. He seems to think that war—aided by antagonism and diversity in religion and language—drove human beings to spread out and occupy the entire globe. Then, through commerce, we will be forced to reconcile and put an end to war. It's as if Kant has in mind a kind of global prodigal son story of separation and reconciliation. The trajectory of this history is driven by natural inclination but in support of a moral end. The diversity of humanity is a kind of necessary evil, making universal monarchy impossible. But the purpose is a universal human community where that difference is wiped away.

utopianism as a kind of *as if* utopianism. It is rational to expect providence to shape the world according to the promise of the moral law, and since we can know what the law commands, we have a good idea what such a future would look like. But the path that nature chooses is likely to surprise us.

The second qualification concerns the content of what is to be perfected, what the student will gain from a good education. It should come as no surprise, after the preceding section, that at least as it concerns the moral domain this will have something to do with the dynamic of humiliation and consciousness of the nobility of the moral law within. There is a worry when it comes to this kind of utopian universal history that the result will be a world of top-down moral bullying, wherein moral subordinates are forced to conform to the standards of moralizing elites. And, as we saw in chapter one, this kind of paternalist threat is likely to conflict with a common sense moral world-view that is suspicious of attempts by those in power to impose their morality on others. Indeed, such pretensions of moral authority are often the targets of those who may be inclined to favor relativism. Relativists would do better, however, to endorse a principle of universal human equality and tolerance if they want to avoid this unfortunate future, and this is exactly what the Kantian moral education ends up promoting. When it comes to the task of educating children, Kant is clear that we should raise them to see that that all human beings are fundamentally equal, in virtue our relation to the moral law.

4.4.2. An Evolutionary Account of Awe

It is helpful to contrast Kant's view with a contemporary treatment to the feeling of sublime humility that Kant links to moral feeling. We needn't necessarily assume that this feeling is a

mysterious point of contact with the Kant's noumenal unconditioned. Dacher Keltner and

Jonathan Haidt describe the feeling of 'awe' in terms of prototypical instances. They see it as a

characterized by "an overpowering and novel sensory experience that causes confusion and

amazement." And they link the experience of awe to transformations in a person's worldview:

"when the confusion lifts, the person is transformed and embraces new values, commands, and

missions."[176] Like Kant's account of the sublime, they associate paradigmatic feelings of awe

with experiences of things that are very large. On their view, the stimulus involved in

experiences of awe are "vast" and require "accommodation."[177] They characterize vastness as

"anything that is experienced as being much larger than the self, or the self's ordinary level of

experience or frame of reference."[178] They also link vastness to power. The concept of

accommodation is derived from 20th Century psychologist Jean Piaget and refers to a "process of

adjusting mental structures that cannot assimilate a new experience."[179] These two aspects of

awe might lead one to expect an account similar to Kant's description of sublimity. When

something's really big (physically or metaphorically), we experience a feeling of awe. This

feeling is inspired by the way in which the object breaks out beyond our ordinary ways of

representing objects, and this requires an adjustment of the structure of representation. For

example, the notion of *eternity,* compared to which one's entire life seems insignificant. This

forces a revision of the ordinary idea of what it means for something 'to take a long time.' It

makes one feel small and is, in a certain sense, humiliating. But Dachner and Haidt do not go in

---

[176] Keltner and Haidt, 2003: 298.

[177] Keltner and Haidt, 2003: 297

[178] Keltner and Haidt, 2003: 303

[179] Keltner and Haidt, 2003: 304.

this direction. Instead, they identify a core version of the emotion for which they can easily

furnish an explanation from evolutionary psychology. They identify all other versions of the

emotion as elaborations of this primordial root. Primordial awe, in their view, is a "response to

cues of social dominance" that "reinforces and justifies social hierarchies by motivating

commitment to the leader."[180] While they do identify a form of awe directed at virtuous people

and those who are very skilled, they maintain that this feeling is derived from the primordial

feeling rooted in the social hierarchy. The core feeling of awe evolved to furnish social

cooperation by leading lower status individuals to submit the authority of higher status

individuals within the dominance hierarchy.

It's hard to imagine a concept more perfectly opposed to Kant's conception of the moral

conscience than this perverse feeling of awed submission to the authority of a dominant human

authority. The dynamic of humiliation and respect that is Kantian moral conscience involves the

agent's consciousness of the moral law within the him or herself. The object of respect is the

moral law within, not the arbitrary authority of another. The humiliation is more analogous to

what religious believers feel when they compare themselves to god than what gym rats feel when

they compare themselves to the guy with better calves. It's a humiliation so deep that all such

earthly comparisons are less than trivial by comparison. In the *Lectures on Pedagogy,* Kant links

conscience to religion. "The law within us is called conscience." But, conversely, "religion is the

law in us, insofar as it receives emphasis form a lawgiver and a judge above us; it is morals

applied to the knowledge of god" (Pedagogy, 9:494-5). The prince—or the largest chimp—

---

[180] Keltner and Haidt, 2003 307-8.

cannot be the author of the law, since the law is grounded in us.[181] Even the "fear of God's power" is a "perverted concept," and Kant recommends that we steer clear of teaching children too much about religion without first having sufficient moral understanding. The worry is that children will develop the idea of God as a powerful authority who rules, by force rather than as the embodiment of moral authority. "If religion is not combined with morality then it becomes nothing more than currying favor" (Pedagogy, 9:494). In opposition to a religion of the latter sort, Kant construes conscience as "the representative of God who has erected a sublime seat above us but also a judge's seat within us" (Pedagogy, 9:495). Insofar as Dachner and Haidt's core theory of awe amounts to a feeling of servility, it can never be identified with moral feeling. *But* if the feeling can be redirected toward a more appropriate object, a kind of moral ideal, then it may be possible to—perhaps by directing our attention to the moral law—awaken the consciousness of morality within. In the previous chapter we discussed Haidt's views on 'sacredness' in the context of a sentimentalist theory of conscience. In *The Happiness Hypothesis* and *The Righteous Mind*, Haidt discusses awe in terms of the moral emotions of 'sacredness' or 'divinity' which he links to the notion of a self that is divided against itself.[182] Unlike Rousseau, Kant thinks that reason and culture elevate humanity infinitely above the rest of creation (Anthropology, 7:127). Whereas it is at best ambiguous in Rousseau, Kant clearly rejects the idea that we are good in the state of nature; we must have the savage disciplined out of us (Pedagogy,

---

[181] From Pedagogy, 9:494: "The human being is contemptible to himself when he is vicious. This is grounded in the human being himself, and it is not a so merely because God has forbidden evil. For it is not necessary that the lawgiver also at the same time be the author [*Urheber*] of the law. Thus a prince can forbid stealing in his land, and yet he could not be called the author of the prohibition of theft. From this the human being learns to understand his good conduct alone makes him worthy of happiness. The divine law must appear at the same time as a law of nature, for it is not arbitrary. That is why to all morality [*Moralität*] there belongs religion."

[182] Haidt, 2006: 200-206, Haidt, 2012: 170-77.

9:441-2). So, if we suppose that Keltner and Haidt are correct about the origin of core awe, this is not inconsistent with the idea that—through culture, religion, moral education etc.—the core emotion has been developed in such a way that it awakens and enlivens the consciousness of the moral law and the fundamental truth that all humans are of fundamentally equal worth.

Kantian moral education is thus geared toward the project (similar to Rousseau's project in *Emile*) of re-conceiving self-evaluation in terms that are appropriate for a community of equals. Children may not shame one another. Pride should be avoided (Pedagogy, 9:491). Children shouldn't be allowed to notice social classes or inequalities among human beings. They can't give orders to the servants (Pedagogy, 9:498). This is all because Kant shares Rousseau's belief that children only come to believe in inequality if they are taught. He's not naïve. Anyone who's spent even a little time with kids knows that they can be perceptive. Kant is aware that the wicked lesson of inequality can be taught to them subtly, without the adults noticing. So, children have to be protected from sliding into feelings of envy when they see other children being praised. Praise leads to comparison and self-evaluation. If they evaluate themselves in relation to others. "Envy is aroused when one points out to a child to value itself according to the value of others" (Pedagogy, 9:491). When children begin to value themselves in terms of how they size up to others, and what others think of them, the result is a dialectic of pride and envy:

> It is quite wrong to let humility consist in valuing oneself less than others. – "See how such and such child behaves!" and so forth. An exclamation of this kind produces a quite ignoble way of thinking. When the human being values his worth according to others, he seeks either to raise himself above others or to diminish the value of the other one. The latter, however, is envy. One then always tries to impute wrong to the other one. For if he were not there, then one also could not be compared with him, and so one would be the best (Pedagogy, 9:491).

If this is simply parenting advice, then it's fine as far as it goes: don't berate your kids by talking them down in comparison to others. Teach them to appreciate the innate value of human beings. But I think the point is broader than that. This passage suggests that the notion of humility that Kant has in mind is antithetical to the hierarchical thinking Keltner and Haidt associate with the core feeling of awe. Humility is not subservience or self-deprecation. It is rather the distinctively moral recognition of the gap between the moral order of things and the natural order of things. Humility is not the comparison of self to others. In Kant's words, it "is actually nothing else than a comparison of one's worth with moral perfection" (Pedagogy, 9:491).

The observations concerning religion and equality that I have highlighted in Kant's pedagogy find an interesting parallel in recent psychological literature on self-compassion. Serena Chen has argued that self-compassion can have a positive impact by increasing the motivation towards self-improvement and decreasing tolerance of our own immoral behavior.[183] In one experiment, self-compassion was observed to be more effective than self-esteem at motivating hard work after an experience of failure. After failing a difficult exam, subjects (students at UC Berkeley) were placed in the self-esteem condition by being reminded of their accomplishments: "If you had difficulty with the test you just took, try not to feel bad about yourself—you must be intelligent if you got into Berkeley." But the subjects in the self-compassion condition were reminded of how struggle is common: "If you had trouble with the test you just took, you're not alone. It's common for students to have difficulty with tests like this." Rather than being reminded of how they stand out relative to others, subjects in the self-compassion condition were reminded that no one is perfect and everyone must struggle to do

---

[183] Breines and Chen, 2012. Wang et. al., 2017.

their best. Subjects in the latter condition showed increased study time and higher scores for the second test.[184] If I am right in emphasizing the role of the moral religion in Kant's pedagogy and contrasting it to the negative influence of pride diagnosed by Rousseau, then the empirical support for the effectiveness of self-compassion lends some to the Kantian view.

In particular, the Kantian approach to awe may be more pedagogically efficacious than the core-awe described by Keltner and Haidt. The struggle for rank may be demotivating for those who lose out, and those who win out may lose whatever incentive there may be once rank is achieved. Recognition of an eternal and universal struggle for improvement has the potential to both console and motivate. This is highly speculative, of course, but the work of Chen and others on self-compassion warrants a degree of optimism.

In addition, we should be bothered by the possibility that hierarchical thinking may come more naturally. The effects of self-compassion show that how we think about what we are doing can influence our action. Finding the right way to think about what we do is a process. Some ways of dealing with challenges will be more effective than others; they may bring what see ourselves as needing to do better into line with what we truly ought to do. It's common to think about Kantian morality in terms of the coherence of our fundamental moral principles. But, as Korsgaard has shown, there is room for discussion as to what such 'contradiction' may imply.[185] If a more teleological interpretation of is correct, it is fitting to suppose that rationality in ethics has less to do with the decision procedure of the FUL than a holistic moral psychology that promotes respect for the moral law and the moral agency that make it possible. Pedagogy would

---

[184] Breines and Chen, 2012: 1137-38.

[185] Korsgaard, 1996b.

thus play a role in brining cognition into line with moral behavior in a way that makes virtue possible and fulfilling.

4.5. Conclusion

In this chapter we have seen that Kant has a distinctive theory of conscience grounded in the consciousness of the moral law. This theory of conscience is capable of providing what Rousseau's cannot: a distinction that separates the moral conscience from mere emotion. Contrary to many first impressions, Kant's moral philosophy is not simply cold rational theorizing. Instead, we have seen that his view furnishes an account of the characteristic moral feeling of humility and respect. This feeling is linked to a feeling of awe or sublime inspiration. In this chapter and the previous one, we have drawn connections between Rousseau's sentimentalist view and the views of Jesse Prinz's and Jonathan Haidt. Like Prinz, Kant and Rousseau both hold that culture can affect human nature in radical ways. Like Haidt, Kant and Rousseau both see something morally significant in moral the feeling of being a divided within oneself. It's worth suggesting that, when he distinguishes the ethics of sacredness from the ethics of autonomy, Haidt may be missing an essential feature of the 'logic' of autonomy. Sacredness involves being a self that is divided against oneself, but autonomy does too. Consciousness of freedom involves a sense of being divided from the person you might wish to be, and the sense of obligation brings with it an immense (unconditional) task. By considering the logical and phenomenal implications of this, Kant is able to identify a characteristically moral feeling for finite rational beings. To compare oneself to the ideal makes one feel small, but this is decidedly different from feeling inferior to another person. In comparison to the holy will, all finite rational

beings find themselves in the same humble boat. Reflection on this fact warrants the recognition that all human beings, in their inward encounter with the fundamental ground of value, are indeed of equal worth.

5.   CONSCIENCE AND SOCIAL CRITIQUE

5.1.   Introduction

This chapter will situate the account of evaluative perception defended in the previous two

chapters within a broader moral theory. The moral theory considered here is a version of the self-

conception strategy.[186] So I will begin by sketching the self-conception strategy. Then I will

defend my preferred version of it in two steps. First, I will distinguish between constitutivist and

perfectionist approaches to the strategy and argue that (with the aid of a reliable theory of

evaluative perception) the latter can avoid what I call the constitutivist's dilemma. The second

step is to show that a theory of evaluative perception that prizes freedom needn't lack moral

content. The key to responding to this worry is to embrace a version of the self-conception

strategy that emphasizes the sociality of agency. The contours of respect are then determined by

social facts about the various identities and relations that we take up in our everyday lives. If

agency is fundamentally social, the self-conception strategy can make use of such facts as

essential elements of an agent's conception of themselves. The resulting view is one according to

which the agent conceives of him or herself in terms of the way in which the fundamental value

of autonomy is manifested through various social practices.[187] An interesting consequence of this

view—surely troubling to some Kantians—is that the universal law formulation (FUL) of the

categorical imperative gets demoted from the supreme principle of morality to a more auxiliary

role. The appropriate role for FUL in this view is as a tool for refocusing moral perception. In

---

[186] While not pejorative, the name "self-conception strategy" comes from David Copp's critique of the approach. Copp, w0o4: 35-39.

[187] We have already seen this sort of strategy at work in Chapter Three, in the response to Jonathan Haidt's cases of moral dumbfounding. See 3.4.

particular, FUL serves to highlight ways in which agents fail to respect others by making exceptions from themselves. Supplemented with a moral sense that sees the value of autonomy, this FUL helps agents see the extent to which they are failing to live up to their ideal conception of self failing to respect the autonomy of others.

In sections 5.2 and 5.3, I argue for for the perfectionist version of the self-conception strategy over the constitutivist version. In 5.4, I consider arguments for the sociality thesis (ST): the view that agency is fundamentally social. If these arguments succeed, then we can be sure that the idea of an isolated agent is a mere abstraction; the view treats one part of agency as if it were the whole. But the arguments don't succeed. Instead, they point toward a compelling picture of agency that they cannot fully establish. In 5.5, I argue that the FUL is a valuable tool in moral philosophy because it serves to highlight another form of abstraction, that of treating one's own interests and agency as the only meaningful elements of one's action. I conclude with some remarks on how the notion of abstraction that I am using as a framework for criticizing the isolated notion of agency and violations of FUL can be useful for moral philosophy

5.2.   The Self-Conception Strategy

The term "*self-conception strategy*" refers to a family of moral views that attempt to ground morality in an agent's ability to form a conception of him or herself. Proponents typically argue there is some basic set of conditions that must be met in order to form an adequate self-conception. Since these conditions serve as the basis for any self-conception that an agent might have, they are said to be the agents deepest moral commitments. Conditions may include: basic self-awareness, autonomy, or consistency in willing, to name a few. The self-conception strategy

is often combined with *constitutivism*, namely the view that being an agent just is (constitutes) adherence to the basic principles of morality. Thus, on what I will call the constitutivist self-conception view (CSC), to conceive of oneself as a an agent one must conceive of oneself as adhering to the basic principles of morality.

The self-conception view, however, needn't just be understood according to the constitutivism of CSC. As an alternative, one might view the self-conception strategy in *perfectionist* terms. That is, one might hold to an *ideal* of what it means to be a good moral agent and consider oneself moral to the extent that they approximate that ideal. According to this view, the moral standard is set by the ideal, and agents are good to the extent that they can conceive of themselves as living in accordance with it. I will call this the perfectionist self-conception view (PSC). The problem with this sort of view is somewhat obvious. The perfectionist must have an ideal in place, and it's hard to see where they would get this ideal without begging the question in favor of that ideal. However, in the next section (1.3), I will argue that CSC faces an insurmountable obstacle that PSC does not. Then (in 1.4) I will argue that things aren't as bad as they seem for PSC. If we combine the account of moral feeling from the previous two chapters with a robustly social conception of human agency, the non-arbitrary and context specific contours of the moral ideal become more perspicuous.

5.3.   The Constitutivist's Dilemma

The conditions of agency that serve as the moral standard are either the conditions of *mere agency* or the conditions of *genuine agency*. If they are the conditions of mere agency, then CSC runs into the emptiness objection and the rigorism objection. The first of these objections says

that the conditions of mere agency simply cannot furnish the moral content required for an acceptable moral theory, and the second says that, even if we suppose it could furnish moral content, it would reduce all immoral action to the same basic failing. If the conditions of that serve as the moral standard are the conditions of genuine agency, then CSC emptiness and rigorism can be avoided, but CSC is faced with a second dilemma.

Either the conditions of mere agency *do* or *do not* tell us something about what it is to be a genuine agent. If they do not, then it seems that CSC gains nothing from linking the moral standard to agency in the first place. This is because, apropos of agency, the moral standard could be anything. CSC has not defined morality in terms of the conditions of agency; instead, it has invented some independent moral ideal. If the conditions of agency do tell us something about the conditions of genuine agency, the view begins to look like a version of natural law theory, whereby some very basic facts about human nature—in this case, what it is to be an agent— serve as the basis for a theory of the good life. The conditions for the possibility of agency are thought to point toward an ideal picture of what an agent ought to be like. But now the view is starting to sound like PSC, and CSC loses what made it seem so promising in the first place. While the conditions of *mere* agency seem unavoidable, it's up to us to decide whether and why we should be *genuine* agents. Just as much as PSC, a CSC view that identifies the moral standard with the conditions of genuine agency will need to defend an ideal conception of moral agency.

So, the constitutivists dilemma can be summarized like this: CSC must either face up to the emptiness and rigorism objections or embrace PSC. Since I think that the emptiness and rigorism objections are decisive, I favor PSC over CSC. The remainder of this section sketches those objections as they relate to some of the early writings of Christine Korsgaard. Korsgaard's

views are complex and sophisticated, but understood in one way, they reflect perhaps the most

powerful version of CSC. It is important to acknowledge, however, that her work (especially

more recently) reflects perfectionist tendencies as well.[188]

5.3.1. Emptiness

The emptiness objection has a long history, and I will not go into that history here.[189] To illustrate

the problem, it's helpful to consider Kant's own attempts to derive duties from the categorical

imperative. Consider the famous example of the lying promise (**G**, 4:402-3). Supposedly, there is

a contradiction when one attempts to will the universalized maxim: 'whenever one is need of

cash, that person may borrow it from someone else with the promise to repay, despite not

actually intending to repay.' According to Korsgaard's interpretation, willing this is inconsistent

because it involves what she calls a *practical* contradiction.[190] This means that the practical

efficacy of the action is undermined by the universalization of the principle underlying that

action. So, if only one person were to tell a lying promise, it's easy to see how they might

---

[188] This is especially true in her discussion of Plato (e.g. Korsgaard, 1999). My understanding of Korsgaard's constitutivist strain reflects the influence of a collection of essays published in *Philosophia* in 2016: Arruda, 2016; Baiasu, 2016; Bratu and Dittmeyer, 2016; Hanisch, 2016; and Hanisch and Baiasu, 2016. Many of these authors deal with Korsgaard's view directly treat it as an example of constitutivism. If her new is not really constitutivist in this way, it seems that there is a tendency by others to read her work as such. I make no claim to be making a definitive statement concerning her view, and in what follows I will focus on themes running through her work on the FUL (esp., Korsgaard, 1996b, 1999).

[189] Wood cites Gottlob August Tittel's 1786 *Über Herr Kants Moralreform,* along with some passages from Fichte's *Sittenlehre,* as among the earliest examples. See Wood, 1990: 155-6. The objection is most commonly associated with Hegel, (e.g. **GPR**, 2.3). But the most common version of the complaint is expressed in a kind of bookend to Mill's *Utilitarianism* (Mill, 1863: 5-6, 77). There is not sufficient space here to treat these historical instances of the criticism, because the nuances of the particular views are important. For example, it is difficult to divorce Hegel's version of the criticism from the role it plays in the systematic presentation of Hegel's own view. And, as Korsgaard correctly points out, Mill's interpretation of the Universal Law Formulation in *Utilitarianism* reflects an unduly rigid notion of contradiction (Korsgaard, 1996b: 80-81). In what follows, I will consider Korsgaard's own Practical Contradiction Interpretation of the Categorical Imperative and subject it to a version the Empty Formalism Objection inspired by Gibbard (1999).

[190] Korsgaard, 1996b, 1998.

succeed. The lender might suppose that this particular borrower intended to return the money because most borrows intend to repay their debts. Presumably, this one person intends to repay the debt as well. This makes the subjective maxim: 'whenever *I am* in need of cash, *I may…*" practically efficacious. But the practical efficacy of the lying promise would undermined under the conditions created by the universalization of this same maxim: "whenever *any person is* in need of cash, *said person may…*" Korsgaard sometimes speaks of "the world of the universalized maxim," a possible world in which everyone is free to do what the subjective maxim describes. In this possible world, the conditions have changed. If everyone is free to lie in order to gets some easy money, lenders would expect only to lose money through lending and, presumably, wouldn't bother to engage in the practice. Therefore, the attempt to simultaneously will the subjective maxim and the universalization of said maxim involves a contradiction, since the conditions created by the universalization of the maxim would undermine the practical efficacy of the subjective maxim.

This is a profound result. In addition to helping to make sense of Kant's universal law formulation, Korsgaard's practical contradiction interpretation of the categorical imperative highlights an important feature of immoral actions. Sometimes we take advantage of others by manipulating their wills toward our own ends. The liar bypasses the will of the person being lied to. This is evident from the fact that lying wouldn't work (at least, not in the same way) if the person being lied to knew what was going on. In addition, lying takes advantage of the wills of all those people who were honest enough to tell the truth in the first place. Those people created the conditions under which the lie was practically efficacious. In this respect, Korsgaard's

practical contradiction interpretation does a good job shining a light on at least some ways in which our actions can be unfair to others.

The constitutivist version of the self-conception strategy goes further, arguing that the consistency demanded by the categorical imperative is a necessary condition for moral agency. As Korsgaard puts it, "Kant's analysis does not reduce the normative concept to a non-normative one; instead it reduces normative content to normative form."[191] Recognizing one's status as a moral agent requires taking one's reasons for action as authoritative for all rational beings. Insofar as one's reasons for action cannot be endorsed by all rational beings, one is unable to conceive of oneself as an agent.[192] And this kind of universality—i.e. the possibility of being endorsed by all rational beings—is exactly what the categorical imperative is intended to test for. Therefore, violations of the categorical imperative, like lying, are morally wrong. The egoist, by excepting him or herself from such rules, is a kind of non-agent and a "practical solipsist" unable to share the reasons for his or her actions with others.[193]

The empty formalism objection rightly rejects this. Practical consistency in willing is not a necessary condition for being (or regarding oneself as) an agent. Allan Gibbard illustrates this point with the example of an ideally coherent Caligula.[194] If Caligula wills to torture the weak for

---

[191] Korsgaard, 1996c: 66. The argument for this is a detailed discussion of the analysis of obligation in the first section of the *Groundwork*. Korsgaard concludes that "only those [maxims] to which my own will commits me…are intrinsically normative." And, since only intrinsically normative principles are capable of obligating us, "right actions are those whose maxims have law-like form, which is the form of normatively itself" (65-66). The objection being developed here is both that the results of such a view are unacceptable *and* that the analysis of obligation fails to provide what the view seems to promise. The reduction of normative content to normative form faces the same problem as other normative reductions (say, to pleasure) because it would still rest on a normative principle—namely, that we ought to prefer formally consistent normative principles. Once such a principle is challenged, the constitutivist faces the same basic problem.

[192] See also: Westphal, 2016; O'Neill 2000.

[193] Copp, 2004: 35-39; Korsgaard 1996a: 143.

[194] Gibbard, 1999. Street, 2004 also uses this example to provide a helpful taxonomy of realist and anti-realist views.

fun, there is no obvious reason why this should undermine his status as a *mere* agent. First of all, it isn't even obvious that the categorical imperative would find Caligula's maxim inconsistent; if everyone were permitted to torture the weak, Caligula might still derive great fun from doing so himself. Furthermore, nothing about agency itself seems to commit Caligula to universalizing the reasons behind his action to include the interests of others. Even if it can be shown that practical consistency in willing requires abstaining from immoral acts, this doesn't seem to show that those who act immorally fail to be agents. They fail to be good agents. Consider Gibbard's dismissal of the "practical solipsism" charge:

> Why must reasons they could claim to each other to have all be reasons they must think they have in common? "I have my reasons and you have yours" is a statement in public language, however unsavory. "Nagel characterized the egoist as a practical solipsist and of course he is right," Korsgaard pronounces. But he isn't right if he mans that an egoist can't explain himself if he likes.[195]

It seems evident that the ability to act according to an evil maxim is just as much a feature of agency as the ability to act according to a moral one, just as the ability to articulate self-serving reasons for an action is just as much a feature of public language as the ability to give universal ones. Reasons for action can be perfectly intelligible and deeply immoral at the same time. Thus, it's hard to see how the conditions of mere agency can be the criteria of morality.

5.3.2. Rigorism

Even if we suppose that the constitutivist can fill the theory out with moral principles, a problem remains. Richard Joyce cites a passage from *The Sources of Normativity* that suggests that this may be what Korsgaard is up to. To commit a moral transgression is:

---

[195] Gibbard, 1999: 163. Quotation from Korsgaard, 1996a: 143.

> …to no longer be able to think of yourself under the description under which you value yourself and find your life to be worth living… it is to be for all practical purposes dead or worse than dead.[196]

Joyce finds this language to be wildly out of whack. It just isn't an accurate descriptions of what's wrong with most ordinary moral transgressions. He uses the example of swiping my neighbor's newspaper from a hotel hallway.[197] This minor transgression does little, if anything, to undermine an agent's deepest conception of him or herself; for most people, such a transgression would have nothing to do with why they find "life to be worth living." The problem here seems to be that the self-conception strategy—understanding agency in abstract universalistic terms—provides the same explanation for both the trivial and the heinous moral infractions.

One might be tempted to allow the details of the case to take care of this problem, and this strategy might work in some cases. For example, the slaveholder's conception of himself as an honorable gentleman is much more important to his self-conception than Joyce's newspaper-pincher's conception of him or herself as a good hotel guest. And this is fitting since the crime of slavery is far more severe than the latter. However, the constitutivist version of the self-conception strategy still cuts against common sense when we consider that the practical contradiction interpretation of the categorical imperative was supposed to identify the conditions of agency—and, by extension, the conditions of conceiving oneself as an agent. Even trivial instances of dishonesty (lying to spare a someone's feelings) are frustrated by universalization,

---

[196] Korsgaard, 1996: 102. Quoted in Joyce, 2006: 197.

[197] It may even be a stretch to call this a *moral* transgression at all, but the practical contradiction interpretation of the FUL makes it hard to make that judgment. Indeed, swiping my neighbor's paper frustrates my purposes when universalized, since it would spoil the system that made the paper available for stealing in the first place.

but as we have seen, quite monstrous crimes (torturing the week for fun) can be universalized without difficulty. It is implausible to think that *the torturer* and *the conflicted friend* who lies to spare a friend's feelings are both unable to think of themselves under the description according which they find their lives worth living. And if they are unable to do so *for the same reason*, there is almost certainly something wrong with that reason.

Both the emptiness objection and the rigorism objection indicate an important problem with the constitutivist approach to the self-conception strategy. By making practical consistency the criterion for agency and agency the criterion for morality, the strategy leads to an implausible picture of morality.

5.3.3. Constitutions or Perfections?

It seems to me that the only way out of the rigorism and emptiness objections is for CSC is to invoke the notion of genuine agency. That is, the moral standard is not furnished by the conditions of mere agency but, rather, the morally thick conditions of genuine agency. I have already suggested that this is a step toward abandoning constitutivism in favor of what I'm calling the perfectionist view. But why this should be such a bad thing?   The main obstacle to the PSC style view is that the appeal to the moral ideal seems like question-begging. Why choose this notion of the good life rather than that one? Proponents of PSC can respond in two moves. First, they can appeal to a source of moral knowledge that is not arbitrary. For that we have the approach to evaluative perception outlined in the previous chapters. Second, the must put this source to work in order to illustrate how it can furnish non-arbitrary moral ideal. This second move is important for us because the conception of the moral sense outlined in the previous chapter is very thin. It is hard to see how a moral ideal follows out of the mere respect for

autonomy. Put another way, the theory of evaluative perception outlined in the previous chapter faces its own version of the empty formalism objection: How can mere respect for autonomy furnish concrete moral principles? The answer is by appealing to what I will call, "the sociality thesis."

5.4.   Abstraction No.1: The Isolated Agent?

In this section, I will outline the sociality thesis and consider some arguments in support of it. The four arguments I will consider are: (1) the epistemological argument; (2) the moral argument; (3) an argument from empirical psychology; and (4) the argument from cognitive stability. Unfortunately, all four arguments fail to prove the sociality thesis. The first two have critical value insofar as they show that denying the sociality thesis may come at a cost. Arguments three and four have value as illustrations insofar they help show that the sociality thesis fits well with certain observations about the human experience.

The sociality thesis may be expressed:

(ST) Individual agency always already contains social conditions that make it possible.

Implicitly contained in the thesis is the notion that agency is made possible (in part) by social conditions (sociality) So, ST entails the social necessity thesis:

(SN) Sociality is necessary for individual agency.

And, naturally, the plausibility of ST and SN will depend heavily on how one understands the central terms: "agency" and "sociality." So, if someone asks whether Tom Hanks's character in the 2000 movie *Cast Away* is an agent, the response from SN must be that *no, he is not*: in complete isolation, there exists no agency worthy of the name. That's why the Hanks's character

can't hold himself together and starts making up imaginary friends, like 'Wilson.' Sociality is

necessary for agency, and he loses the latter as he loses the former. We can qualify this somewhat

by saying that sociality persists in the (perhaps in a deficient way) insofar as the the character

talks to himself; continues speaking English; and thinks about his past. One consequence of ST

and SN is that the contours of sociality can be hard to draw clearly, but if SN is true, you can't

have agency without sociality.

If true, the sociality thesis would be a great benefit for PSC. IF sociality is necessary for

agency, then conceiving of oneself as an agent—and *a fortiori* conceiving of oneself according to

the ideal of a perfect agent—requires invoking elements of one's social context. In the second

argument for ST we will see that Kwame Anthony Appiah's notion of an "honor world" provides

additional material with which to imagine oneself as the kind of agent one wants to be. On his

view, moral revolutions take place when honor worlds break down and practices that were once

viewed as honorable come to be viewed as disgraceful.[198] He uses the example of dueling and its

correlated ideal of "the gentleman." To be a gentleman, a man needed to be willing to defend his

honor in the right kind of way. The duel had a complex system of rules. It would have been silly

to duel with a woman or a poor person. Appiah recounts how the Prime Minister, the Duke of

Wellington, made a joke of himself and his office by engaging in a duel with the Duke of

Winchilsea.[199] The duel, as part of an out-dated aristocratic honor code, was unfit for the chief

representative of the citizens in a democracy. The class structure that once served as the context

---

[198] Appiah, 2010: 161-62.

[199] Appiah, 2010: 3-26.

for the gentleman's honor code had collapsed into something else. In acting like a Duke,

Wellington made a fool of the Prime Minister.

Agency involves a kind of self-understanding, and sociality mediates that self

understanding, giving the agent the concepts according to which he or she can view a life as

valuable. From this it follows that *the significance of our actions is not entirely up to us*. We are

always already engaged in a network of relations that give our actions their significance, and this

places a limit on the self-conception strategy. Individuals cannot view themselves as

approximating an ideal unless they can understand themselves as situated within a network of

social relations. This network of social relations determines what it means, in any given case, to

treat an agent with respect. If ST is true, then the notion of an isolated agent is merely an

abstraction.

5.4.1. Four Arguments for the Sociality of Agency

Here I will present four arguments for the sociality thesis. They all fail to prove the thesis;

instead, they point toward its initial plausibility. With additional support, they could form part of

an argument by inference to the best explanation.

5.4.1.1. An Epistemic Argument

Kenneth Westphal has argued that Hegel's thesis of mutual recognition can overcome problems

in epistemology.[200] Specifically, he thinks that by embracing pragmatic fallibilism and

constructive self-criticism, we can avoid Sextus's famous Dilemma of the Criterion. This well-

known dilemma undercuts foundationalism by calling its criterion into question. The criterion if

---

[200] Westphal, 2009.

truth is either assumed or dependent on something else. In the former case, it is question-

begging. In the latter case, there is either an infinite regress of criteria or it bottoms out in the

same question-begging assumption. But by embracing pragmatic fallibilism and constructive

self-criticism, we can proceed rationally from where we're at, without supposing that we've

found the ultimate or most fundamental criterion.[201]

In this context, Hegel's thesis of mutual recognition, and his view that "individual

rational judgment… is socially and historically based," derives additional plausibility.[202] Our

epistemic situation is one of fallibility and self-criticism. Thus, it becomes essential to view

others as as "equally fallible and equally competent to judge."[203] Acknowledging this involves

acknowledging that we need to take into account the assessments and perspectives of others in

order to rationally scrutinize our own judgment. Rationality involves being answerable to what

others think. Thus, to bring it back to ST, we might argue that, since rationality is a condition of

agency, and that since rationality requires being answerable to others, agency requires sociality.

I have a great deal of sympathy for Westphal's account. I think he is right that self-

criticism in the context of something like pragmatic fallibilism can provide a meaningful basis

for revising judgment, that is consistent with objectivity and even realism about the objects of

---

[201] "In sum, the key to Hegel's reply to the trope of circularity is to show that, because we are capable of constructive self-criticism, when assessing or reassessing any piece of justificatory reasoning, by reviewing its basic evidence, principles of inference, and its use of these, we can revise, replace, or reaffirm as need be any component and any link among components within the justificatory reasoning in question." From Westlphal, 2009: 771.

[202] "This paper aims to show that Hegel is correct, that individual rational judgement—of the kind required for rational justification, whether in cognition or morals (broadly construed)—is socially and historically based, although the bases of rational judgment identified by Hegel are consistent with realism about the objects of knowledge and with strict objectivity about moral principles." Westphal, 2009: 754.

[203] Westphal, 2009: 762. Here he is characterizing Hegel's thesis of mutual recognition, as it is expressed in **PdG**, p. 666-71.

those judgments.[204] I'm also sensitive to the idea that completely disregarding the views of others is incompatible with rational judgment. Still, there are two major problems with this type of proof. First, it's unlikely to persuade epistemologists who are unperturbed by the Dilemma of the Criterion. And I'm not sure why foundationalists couldn't just leave that problem unresolved, proudly 'begging the question' on something that they take to be self-evident.[205] If circularity is merely a trope, then all the less reason to worry. Second, even if fallibilism is true, the sociality thesis still may be false. One can be a rational fallibilist while worrying little about the opinions of others. An additional premise is needed. Perhaps it is by taking the opinions of others and relevant social facts into account that we make our best judgments.

Something like this might be true. Individualistic foundationalists, like Descartes, certainly have a reputation for working in isolation and demanding certainty. I'm inclined to agree that there's something about their overall approach that lends itself to problems like Sextus's dilemma. But the Hegelian approach isn't to isolate the necessary and sufficient conditions of a view like individualism. Instead, it's to identify the motivations and blindspots of an overall picture of the world. Through a strategy that works less like an argument than an illustration, what the view really needs becomes perspicuous. The epistemic argument for ST is understood better as an illustration than as a demonstration. When we think about what we want from a theory of knowledge and how that might play out in the real world, it really does seem as though rational judgment involves being answerable to others, not as epistemic peers but as sources of consideration. The biologist needn't defer to the accountant in biological matters, but

---

[204] See notes 15 and 16 above.

[205] I'm thinking here of the self-evidence response to the regress argument. See 1.5.

in considering the views of other experts—rather than novices—the biologist reveals a genuine interest in social facts as part of the process of rational judgment.

5.4.1.2. A Moral-Epistemic Argument

Like the previous argument, this one is critical. Instead of providing a positive proof of ST, the strategy is to identify troubling weaknesses in more individualistic views. According to this argument, denying ST leads to a broadly "positivist" worldview. Appiah argues that, by rejecting positivism, we can escape moral relativism.[206] And, according to him, the initial mistake that makes positivism seem plausible is the denial of ST.[207] We can summarize the strategy like this: individualism is the denial of ST, and individualism leads to positivism which, in turn, leads to relativism. Therefore, since relativism is false, ST is true.

The first thing to say about this strategy is that it assumes what needs to be proven. That is, it can't show relativism is false; it depends on the assumption that relativism is false. But Appiah's version of the argument shouldn't be understood as a deductive demonstration. Instead, it's meant as an illustration of a more compelling alternative. Individualism doesn't entail positivism like being bachelor entails being unmarried. It just seems to *fit well* with positivism.[208] The positivist's view depends on the Humean belief/desire psychology that, arguably, makes the most sense when we consider individuals in isolation. The central moral question becomes one of personal motivation, and Appiah's move is to deny that personal motivation is what's really

---

[206] Appiah, 2006:

[207] "The deepest problem with Positivism, however, is not in its conclusions. It is in its starting point. I began, as I think one must if one is to make the Positivist story believable, with a single person, acting on her own beliefs and desires. Starting from there, one has to give an account of values that begins with what it is for me—this single person—to regard something as valuable. But to understand how values work, you must see them not as guiding us as individuals on our own but as guiding people who are trying to share their lives." Appiah, 2006: 27.

[208] See note 21.

important. What Appiah's calling positivism is really a general way of looking at moral

psychology that assumes that the most important thing for ethics is to persuade the individual

agent by appealing only to principles that he or she already accepts. The key to his response is to

reject this move and recast moral reasoning as a kind of conversation.[209] What becomes most

important to ethical philosophy, then, is the collective process of "learning to align our responses

to the world."[210]

As with Westphal, the truly compelling part of Appiah's argument is not so much

demonstration but illustration. His broadly externalist view seems to reflect a more accurate

picture of how we engage in moral deliberation. Unfortunately, the closest thing we get to a

proof of the view is the critical analysis of another view (positivism) that is thought to harmonize

with the denial of ST. To the extent that there is a positive proof of his externalist view at all, it's

more of an illustration.

Since it is relevant to my project here, I would like to say a bit more about that

illustration in the remainder of this subsection. Understanding Appiah's account may help to

show what the a moral view grounded in the feeling of respect for autonomy can gain from

external social considerations. According to Appiah, when we engage in moral conversation we

take advantage of systems of meaning that are not merely dependent on what's in our own heads.

"You can talk about elm trees, even if you personally couldn't tell an elm from a beech."[211] And

he extends this thesis to thick moral predicates like kindness.

---

[209] Westphal (2016) makes a similar move when he rejects "justificatory internalism."

[210] Appiah, 2006: 29.

[211] Appiah, 2006: 27.

> The concept of kindness or cruelty enshrines a consensus… Like all vocabulary, evaluative language is primarily a tool we use to talk to one another.[212]

The moral deliberation is not primarily about an individual thinking through his or her subjective desires and beliefs; it is about a collective conversation that humanity is constantly engaging in. One can't simply assert that kindness is bad and cruelty good. Such a person might be justly accused of misunderstanding the terms. He highlights the value of stories people all over the world readily understand and can meaningfully assess in terms of their moral significance and the virtues and vices of the characters. "Evaluating stories together is one of the central human ways of learning to align our responses to the world."[213]

On his view, major changes in the moral conversation rarely the result from arguments, logic, or reason. "The reasons that we exchange in our conversations will seldom do much to persuade others who do not already accept our fundamental evaluative judgments."[214] Using the example of feminism and the movement of women into the workforce, Appiah maintains that this perspectival shift had less to do with arguments than the fact that people "our getting used to the new ways of doing things."[215] While it is surely possible to revise some of our moral beliefs on the basis of inference from other commitments, significant ethical shifts usually involve changes of habit and perspective. His book on moral revolutions extends this line of reasoning to four historical examples: three from history—the end of dueling in England; the end of foot-

---

[212] Appiah, 2006: 28.

[213] Appiah, 2006: 29.

[214] Appiah, 2006: 72.

[215] Appiah, 2006: 76-77. "The arguments that kept the old pattern in place were not—to but it mildly—terribly good. If the *reasons* for the old sexist way of doing things had been the problem, then the women's movement could have been done with in a couple of weeks… Arguments mattered for the women who made the women's movement and the men who responded to them. This I do not mean to deny. But their greatest achievement was to change our habits."

binding in China; and the end of the transatlantic slave trade—and one that he hopes to see in the not-to-distant future—an end to honor killings among the Pashtun populations of Afghanistan and Pakistan. He identifies some common features in each of the historical cases. For instance, each of the moral revolutions involved a shift in "codes of honor." Sometimes he calls this a persons "honor world." And in each case, "the code of honor faced moral and religious challenged long before the revolution." As with his discussion of the women's movement, moral revolutions do not seem to be motivated by moral or religious arguments but a shift in perspective. "At the end of the revolution, honor was successfully recruited to the side of morality."[216] What was once honorable comes to be viewed as disgraceful.

Appiah's account suggests that this is what happened to dueling.[217] He describes to the case of the Duke of Wellington who engaged in a duel in 1829, while he was England's Prime Minister. Motivated by a scandalous accusation to defend his honor, the Duke felt forced to challenge the Earl of Winchilsea to a duel. Due to the prominent political position of the parties, the duel attracted the attention of the broader public and raised some uncomfortable questions about the nature of the "gentlemanly" honor that the participants had hoped to defend. In particular, the honor of a "gentlemen" obviously wasn't extended to women or to men of the lower classes, yet the purpose of the duel was to defend the gentleman's equality of status with a gentleman who had slandered him. So, when it was viewed from the perspective of a broader society with members of the lower classes now asserting an equality of their own, the essentially

---

[216] Appiah, 2010: 161-62.

[217] Appiah, 2010: 3-26.

exclusive and hierarchical notion of the "gentleman" runs into conflict with it's own appeals to

the norm of equality among gentlemen.

> In the world of honor, the equality of gentlemen, displayed in the duel, declared their shared superiority to the common people. Deny that and the whole scheme falls apart.[218]

The result was ridiculous and seems to have embarrassed all parties involved.[219] The duel was no

longer able to do what it was supposed to do—reaffirm the status of the nobility. He quotes a

passage from Charles Greville's *Memoirs* that reflects the perplexing state that the honor world

of the duel was in at that point:

> I think the Duke ought not to have challenged him; it was very juvenile, and he stands in too high a position, and his life is so much *publica cura* that he should have treated him and his letter with the contempt they merited… [but] it is impossible not to admire the high spirit which disdained to shelter itself behind the immunities of his great character and station, and the simplicity, and almost humility, which made him at once descend to the level of Lord Winchelsea, when he might, without subjecting himself to any imputation derogatory to his honor, have assumed a tone of lofty superiority and treated him as unworthy of notice. Still it was beneath his dignity; it lowered him, and was more or less ridiculous.[220]

Greville's twists and turns in this passage reflect his confusion as the honor world of the

gentleman was falling apart. He remained committed to this code of honor, insofar as to admire

the man who risks death in defense of his honor as a gentleman. According to this standard, the

Duke had to do it. But as Prime Minister his honor didn't come from his status *as a gentleman*

but *as a representative in a democracy*. In that role, having the opposition say bad things about

you is just part of the job. From the perspective of this honor world, the Prime Minister lowered

---

[218] Appiah, 2010: 42.

[219] Appiah reprints a political cartoon from the time depicting the Duke's duel (which was held near the Battersea Bridge) under a banner reading, "Send them to Battersea to be Cute for the Simples."

[220] Quoted in Appiah, 2010: 41.

himself, because the status of the Prime Minister (unlike that of the gentleman) comes from the special function that representatives play in a democracy—that of translating the will of the constituents into policy. The personal honor of a gentleman becomes unimportant. Rather than acting like gentleman on equal footing with other gentlemen, he was acting like some guy who had too many drinks at the bar and got into a fistfight over politics.

In the case of dueling, a shift in the social structure reshaped the honor world and gave respect a different meaning. Respect for autonomy was a good thing in the heyday fo the duel, and insofar as dueling reflected such respect it was not without value. But what counts as respect for others seems, to a large extent, socially determined. It is no contradiction to say that the very same deed may reflect respect for autonomy in one context and the opposite of respect of that in another. Appiah concludes his book on moral revolutions with a discussion of Kant. Appiah goes further than just saying that honor can be a powerful moral motive—assuming of course that the relevant honor code is consistent with duty. He certainly thinks this, but after a discussion of Kant's notion of autonomy, he compares our sense of honor to Kant's fact of reason:

> I am arguing, against Kant, that honor is another of the calls on us made by reason; it is a call that depends on our recognition of the many different standards presupposed by those codes of honor. And when those standards make sense to us—when we inhabit the same honor world—we understand as well that those who meet them deserve our respect. Sometimes, as we have seen, the standard will be morality. Often, however, it will not.[221]

Here we see Appiah developing more fully the idea only hinted at in *Cosmopolitanism*, that morality and moral language is about a kind of conversation with others. It is in engaging with those others that the demands of reason become explicit and are imposed on use from the

---

[221] Appiah, 2010: 183.

outside. What it means to respect someone is a function of conditions within what he calls our honor world. It makes perfect sense that this would remain indeterminate when we limit our conversation to the constitutivist's conditions for the mere possibility of agency. Respect for autonomy is a demand of reason, but this respect is meaningless in the abstract. What's needed is social engagement in the context of something like what Appiah calls an "honor world."

5.4.1.3. An Argument from Empirical Psychology

The third argument is a positive one. Recent work in empirical psychology has been points in an externalist direction. We saw this in our discussion of the role of expert intuition and external cues in the context of evaluative perception.[222] If Kahneman and Klein are right that expert intuition requires a high validity *external* environment in which to operate reliably, the reliability of some of our expert intuitions may be thought to lend support an externalist view. Perhaps, as Sloman and Fernbach put it, the mind is not like a computer processing stored information: very little of our knowledge is stored in the head, and we are constantly responding to the external world and picking up bits of information from what we see.

We saw that this sort of explanation does a better job accounting for everyday instances of expert intuition, like catching a fly ball. If this externalist thesis is true of our everyday experiences of the world, it stands to reason that social elements of our environment may shape our experience as well. Instances of prejudice provide obvious examples because their distastefulness makes them especially striking: a man sees a woman in a fancy coat and supposes that she's shallow. But even the fact that we understand how a coat could register as 'fancy'

---

[222] References in this section can be found in section 1.2.2.

suggests that we process social cues at a deeper level—perhaps unconsciously, as we process the relative angle of the fly-ball.

If sociality influences our lived experience at a deep level, it follows that a realistic self-conception will need to include social elements. So, there is some empirical support for the sociality thesis. But there are many ifs in this argument. Externalism would have to go pretty deep to establish SN. To be sure, it's hard to imagine an actual person getting around in the world without relying heavily on social cues, but is this enough to show that sociality is necessary for agency? I'm not sure, but the following argument pushes this question a bit further.

5.4.1.4. An Argument from Cognitive Stability

The argument from cognitive stability rests on the premise that sociality is needed to provide consciousness with stability. On this view, being able to hold our thoughts together is necessary for agency, and sociality is needed to help hold our thoughts together. "Holding thoughts together" is a necessarily vague notion, but it is reflected in everyday phrases, like: "pulling yourself together," "cracking up," and "breaking down." That sociality is required for this seems to be the intuition behind the idea that Tom Hanks's character in *Cast Away* loses his agency.[223]

If sociality is required to hold thought together, then we might say that sociality provides fixed points of orientation that allow the agent view him or herself as part of a world. Keeping it together may already involve a limited notion of self-conception. We may have to suppose that self-conception and agency go hand in hand. That is, conceiving of oneself as an agent is not just the basis of a normative theory but a genuine precondition for agency itself. The two ideas tend

---

[223] Gilles Deleuze (1984) discusses a similar sort of crack-up in an essay on Michel Tournier's, *Friday, or The Other Island*. Deleuze obviously does not share the view reflected in the argument under discussion here, but one may be drawn to the (perhaps, naïve) interpretation—that in the absence of others, Friday loses his agency.

to feed off each other—I am an agent because I conceive of myself as such, and I it is because I do so that I can conceive of myself as an agent. The argument from cognitive stability maintains that sociality is an essential element of how one conceives of oneself and that self-conception is essential for agency. Notice, however, that this needn't put us in the same position as the constitutivist—grounding the self-conception strategy in what constitutes agency—because what grounds the preference for genuine agency (understood as a kind of agency capable of keeping it together) is the feeling of respect for rational agency. The perfectionist version of the self-conception strategy can treat the ideal of the genuine agent as the moral standard because it inspires respect.

Like the first two arguments, the argument from cognitive stability gains some support when we consider its opposite. But, Tom Hanks aside, there aren't many examples of completely isolated agents—the problem with hermits is that they rarely come to visit. It's worth considering some tragic cases of extreme short-term memory loss caused by Korsokoff's syndrome documented by Oliver Sacks. Such cases provide an interesting demonstration of the fact that the inability to get a grip on one's external environment undermines one's ability to get a sense of who they are.

Patients with Korsokov's syndrome experience degeneration of the short term memory, but the long term memory remains mostly intact. I say "mostly" because it's not uncommon for Korsokov's patients to experience retrograde amnesia. Oliver Sack's describes these cases with his characteristic sympathy and insight.[224] One such case is a man he refers to as Jimmie G., who could remember his life up to about 1945 but he could not remember anything since, beyond

---

[224] Descriptions of Korsokov's syndrome and the case studies referred to in the coming paragraphs can be found chapters 2 and 12 of Sacks, 1998.

about two minutes. He was sharp and good at puzzles, but if the puzzles took him too long to

solve, he would forget what he was doing. Sacks would test Jimmie's short-term memory by

showing him small objects and then hiding them. He would set his watch on the table and then

cover it with a magazine. After a few minutes Jimmie would forget that the watch was there.

Whenever Sacks came to visit Jimmie he had to reintroduce himself. Believing that it was still

1945, Jimmie thought he was going to turn twenty soon and that his brother—now retired—was

still studying to be an accountant:

> Looking at the grey haired old man before me, I had an impulse for which I have never forgiven myself—it was, or would have been, the height of cruelty had there been any possibility of Jimmie's remembering it.
> "Here" I said, and thrust the mirror toward him. "Look in the mirror and tell me what you see. Is that a nineteen-year-old looking out from the mirror?"
> He suddenly turned ashen and gripped the sides of the chair. "Jesus Christ," he whispered. "Christ, what's going on? What's happening to me? Is this a nightmare? Am I crazy? Is this a joke?"—and he became frantic, panicked.[225]

After that, Sacks walked Jimmie to the window and distracted him by pointing out what a nice

sunny day it was outside. He left and came back a few minutes later. Jimmie had forgotten the

whole thing, and the doctor had to reintroduce himself and start over.

Another patient, named William, used seems to have used speech as a coping mechanism.

William was constantly talking, making up stories about himself. Sacks interprets these stories as

a way to make sense of his constantly changing (disappearing) mental life. "Abysses of amnesia

continually opened up beneath him, but he would bridge them, nimbly, by fluent confabulations

and fictions of all kinds."[226] The stories that William made up held together for as long as he

needed them to; if they were inconsistent with something he had said earlier, it's probably

---

[225] Sacks, 1998: 25.

[226] Sacks, 1998: 109.

because the earlier assertion was entirely forgotten. "Such a patient *must literally make himself (and his world) up at every moment.*[227] Another patient named Stephen R. remembered his neighborhood as it was prior to 1978, but he was unable to recognize anyone at the hospital where he was being treated. So, sometimes his wife would take him back to his old house, where most everything remained just as he'd left it. He would sit in his favorite chair and move competently around the house, eventually forgetting that he no longer lived there. He seems to have felt at home, although he was confused by changes in the neighborhood and the apparently rapid aging of his neighbors. When he returned to the hospital he had a (brief) experience of terror similar to that which Jimmie felt when he looked in the mirror.

Sacks doesn't directly talk about what I'm calling the argument in favor of cognitive stability, but his discussion of these cases reflects a certain bias toward the worldview underlying the argument. Jimmie's most significant contact with the past came from his brother, who sometimes came to visit. From Sacks's description, Jimmie's brother seems to have had a bit of a baby-face that hadn't changed much over the years, and Jimmie seemed to truly recognize him— even if puzzled by how fast he seemed to be aging. According to Sacks, Jimmie was at his best taking communion in the hospital chapel. He also did well when they were able to put him to work typing and tending the hospital garden. But Sacks reveals his disappointment in what he perceives as the shallowness of Jimmie's experience. After getting Jimmie to write his thoughts in a diary, to help him remember things, he laments that Jimmie's observations "never touched

---

[227] Sacks, 1998: 110.

the depths."[228] But in the chapel, Jimmie seemed to reveal an "intensity and steadiness of attention" and "a perfect alignment of his spirit with the spirit of the Mass."[229] Sacks concludes:

> Perhaps there is a philosophical as well as clinical lesson here: that in Korsakov's, or dementia, or other such catastrophes, however great the organic damage and Humean dissolution there remains the undiminished possibility of reintegration by art, by communion, by touching the human spirit: and this can be preserved in what seems at first a hopeless state of neurological devastation.[230]

We can set aside the swipe at Hume. What's important here is significance of instances of "intensity and steadiness of attention" for a being whose phenomenal experience of the world is so constantly and fundamentally fractured. What makes it so difficult for Jimmie, and others with severe short-term memory disfunction, to hold things together is not the absence of an abstract conception of self. What's lacking is a solid grip on the everyday world around them. Without a stable understanding of how things are *out there*, they are unable to conceive of themselves in a more robust, substantial and socially connected sense. William's efforts to patch his world up with narrative are indicative of a more everyday human struggle to keep it together in a world in flux. But William's world, being in an extreme state of flux, makes a stable mode of existence impossible for him. To borrow a phrase from Rahel Jaeggi, he is unable to "have the world at [his] command."[231] His experience might be regarded as an almost permanent state of alienation.

As powerful, and occasionally disturbing, as these cases are, they fail to support an argument for ST because they fail to isolate sociality as a necessary precondition for agency. Korsakov's patients don't just lack sociality; they lack short term memory. This frustrates their

---

[228] Sacks, 1998: 35.

[229] Sacks, 1998: 37.

[230] Sacks, 1998: 39.

[231] Jaeggi, 2006.

ability to get a grip on the entirety of the outside world, social or otherwise. Stephen R's trips back home provides some support for the importance of at least some external conditions, and we may want to count elements of the social world among these. But the examples don't prove that sociality is necessary for agency.

5.4.2. Is the Isolated Agent A Mere Abstraction?

None of these arguments succeed in proving that the isolated agent is an empty abstraction. Instead, what they do is point to some unfortunate consequences for an individualistic account of agency and illustrate some reasons for thinking that an externalist account of agency makes a good deal of sense. By embracing ST, it seems that the individualist could avoid: (1) the Pyrrhonian dilemma of the criterion and (2) a 'positivist' case for relativism. In addition, the externalist picture reflected in ST, seems to fit well with (3) a potential challenge based on recent empirical psychology and (4) some cases studies and thought experiments concerning the phenomenology of mental life. In the end, while I am sympathetic to this socialized picture of agency, I am skeptical of attempts to prove it. My hope is that these arguments serve to illustrate why such a view is compelling, even though they fail to establish it.

5.5. Abstraction No.2: Making and Exception for Oneself

In this last section I want to say something about how moral deliberation can benefit from the notion of abstraction and suggest a reading of the practical contradiction interpretation of the categorical imperative on that model. Actions found to violate FUL might best be understood as a special case of what Hegel called abstraction. In linking the FUL to the ends in themselves formulation of the categorical imperative, Korsgaard has pointed out that characteristic violations

of FUL (like lying) involve an agent's make an exception for him or herself. The liar takes advantage of the poor suckers who tell the truth and the poor suckers who doesn't know that they're being lied to.[232] One might infer from this that what's really going on in the most characteristic violations of FUL is the following sort of abstraction: the agent's action depends on the cooperation of others, but the agent acts as if the only significant feature of the action was the agent's own personal interest. From the perspective of a conscience that values autonomy, the FUL shines a light on a way in which agents may fail to respect the autonomy of others and, hence, fail to live up to the perfectionist ideal supported by the theory of conscience defended in chapters four and five. In the remaining sections, I will say a bit about how abstraction can be useful in moral theorizing and how the FUL might be situated in our theory as a tool for identifying a specific kind of abstraction that's particularly important to moral deliberation.

5.5.1.   The Notion of Abstraction as a Tool for Moral Philosophy

If ST is true, then any version of the self-conception strategy that fails to take social elements of agency into account is deficient. Such a view would be guilty of what Hegel called "abstraction." Since the version of PSC advocated here takes advantage of the feeling of respect for autonomy, if such respect is manifested through the social elements of agency, PSC is free to draw on those elements. Social practices might be judged in terms of the way in which they manifest respect.

A recurring theme in Hegel's corpus is the notion of abstraction. Following his formulation, one thinks abstractly when they take the part for the whole.

---

[232] Korsgaard, 1998: xi.

This is abstract thinking: to see nothing in the murderer except the abstract fact that he is a murderer, and to annul all other human essence in him with this simple quality" (**WTA**, 463).[233]

With this formulation, Hegel criticizes both the moralist and the apologist. The former sees only the crime and ignores the murderer's fundamental humanity and conditions of his life that lead to the crime. The superficial apologist puts everything on the murder's tragic upbringing, as if he were insufficiently human to determine his own actions.

We have seen some examples of abstraction in the previous chapters, even though I haven't used the term. In chapter three we discussed Jonathan Haidt's cases of moral dumbfounding. In the moral dumbfounding case, we found that the intuitive force of Haidt's example came from the way in which he manipulated the example so that respondents were forced to disregard morally relevant contextual elements of the case. Respondents were forced to consider the action in the abstract, without considering potential morally relevant facts about the individuals broader life experiences.[234] As Appiah's discussion makes clear, customs and social practices often embody a structure of significance—what he called an "honor world." Agents needn't be aware of the elements of this structure in order to express themselves through it. But when agents understand their actions in the abstract, they sometimes miss important elements of the rational order of things. For a simple (and somewhat superficial) example, consider a North American driver who knows that he's supposed to drive on the right hand side of the road. This is the rule almost everywhere, and most of us can get by simply remembering that they are supposed to drive on the right. But anyone who reflects on the matter will see right away that

---

[233] See also, **PG**, 100-102/**PdG**, p. 131; **Enz**, p.511.

[234] Section 3.4.

what's important is not 'driving on the right' but the safe and predicable sharing of the road. A closed-minded driver who goes to the United Kingdom and insists on driving on the right will have missed the point of the rule altogether. He will have taken the shallow and specific rule for the whole and ignored the deeper principle. Like other instances of abstraction, one takes the part (a specific principle of action) for the whole (the principle's purpose in its broader context). *If ST is true, then the isolated agent is an abstraction for essentially the same reason.*

This embedded conception of moral rules sets the Hegelian approach apart. We can see this in his discussion of intention and insight (**Enz**, p.505-12; **GPR**, 119-41), which makes constitutes a crucial move in his transition from "morality" to "ethical life." Here he distinguishes "the right of intention" from "the right of well-being." Note that he doesn't make the move—commonly associated with Kantian ethics—of ignoring the latter and making morality all about intention. He rejects the view that says only intentions (and never consequences) matter. The right of well-being demands that one's action promotes the good, and the right of intention demands that that we act on the belief in the rightness of what we are doing.[235] Both the right of intention and the right of well-being constitute essential elements of morality, but alone, each of these is a mere abstraction. Kant regards the good-will as the necessary condition of one's worthiness to be happy, making happiness seem accidental (**G**,

---

[235] Dean Moyer (2014) puts this nicely by characterizing Hegelian autonomy in terms of non-detachment: "Hegel's attacks on subjective conviction are not attacks on the importance of individual belief in determining duty per se, but are rather attacks on the *detachment* of belief, as one of the components of ethical action, from the normative status of the complex whole" (82). Later on, this allows Moyer to plausibly defend Hegel from the charge of totalitarianism. "Hegel consistently asserts that the state has the higher right than the individual's moral reflection, yet full-fledged action within the state also requires my belief, not my mere obedience to authority. To act simply because the state commanded would also be to detach one part of the complex of objective duty. The action is a duty because the individual believes that it is right *and* the state requires it. If the state must regularly assert its right *against* the agent's belief, the state is not living up to Hegel's idea of freedom… Hegel maintains that the human will always contains a moment of possible difference between the particular and the universal" (107-8).

4:393).[236] If it's taken to mean that intention is the whole of morality, then such a view takes the part (admittedly and essential part) of morality to be the whole and misses the way in which intention belongs to a network of meanings and practice that extend beyond the actor's subjective understanding. The meaning of an agent's action is not fixed by the agent him or herself.

The agent's *intention* must be properly embedded within its broader network of meanings and practices in order to realize its intended effect. So, while intention and consequence are conceptually distinct, they can't be so sharply distinguished in practice. Hegel contrasts *intention* [*Absicht*] with *insight* [*Einsicht*]. The shoe-gazing individualist focuses on the pure intention of an action—one *looks away* from the embedded meaning of the action, which is determined by its social context. Insight into the good is gained by understanding the social structures of meaning that give an action its significance. Abstracted from this network of significance, the formal principle of an action may be the same whether the action is good or evil. Viewing morality this way allows us to see why dueling fell out of fashion. In the context of a nobility distinguishing itself from the lower classes, it was an honorable way of asserting one's status relative to other gentlemen. In the context of a representative democracy, it became degrading and ridiculous. We cannot fully understand an action if we limit ourselves to the formal analysis of intentions. The meaning of an action depends on elements beyond the subjective intentions of the actor. This is how we might understand what Appiah's claim "that honor is another of the calls on us made by reason."[237]

---

[236] Kant's actual view is more nuanced. See 4.4.

[237] Appiah, 2010: 183. See note 35.

5.5.2. The Value of the Universal Law Formulation of the Categorical Imperative

This is also how I think we should understand Kant's universal law formulation of the categorical imperative (FUL). The FUL does a better job identifying instances of institutional injustice than instances of senseless violence. I am persuaded by examples like Allen Gibbard's notion of an ideally coherent Caligula that the FUL is not a good candidate for a fundamental normative principle. But this doesn't mean that the perfectionist version of the self-conception strategy being advocated here has no use for it. The FUL is a powerful tool for focusing our attention on a particular kind of abstraction that is especially relevant to respect for autonomy: making a special exception for oneself.

The view being advocated here is better than the constitutivist version of the self-conception strategy because it avoids the constitutivist's dilemma by boldly embracing perfectionism. And it is better than individualistic perfectionists views because, by making sociality central to agency, requirements of respect are seen to follow from an individuals social roles. In the real-world agents actually are committed to a particular conceptions of themselves. An individual's self-conception involves values, identities, and honor worlds that make up their everyday experience. The duel reflected a commitment to the idea of "gentleman." But the conception of gentlemanliness failed to make good on its promise. What made this self-conception appear valuable was the way in which it supposedly reflected the dignity and inner worth of the gentleman, and in that respect, there is something worthwhile about it. But by excluding the majority of the population, it proved to be little more than haughty posturing. When perspectives shifted to allow assessment by the broader population, it was rightly

subjected to ridicule.[238] Following Westphal, we can say that those who distinguished themselves as 'gentlemen' failed to acknowledge their fallibility and equality with those they excluded from the class. When the exclusion could no longer hold, the practice was seen for what it was. The mistake was not in the logic of *possible* agency; one can easily conceive of oneself in this way. But in taking social status as a substitute for inner worth, they mistook the part for the whole. What has worth is to be honored, but honor is not a substitute for genuine worth. With a perspectival shift, the substitute for worth was seen for what it was and became laughable.

By focusing attention on the inner worth and dignity of autonomous agents, we probably won't find a set of conditions for the possibility of agency, but by looking at social practices in order to determine whether and to what extent they reflect opportunities for genuine agency, we may find self-conceptions and opportunities for agency. Korsgaard's practical contradiction interpretation of the categorical imperative, along with the Hegelian notion of abstraction, can help refocus our attention on the ways in which actions and social practices may fail to justice to the wills of others. One's own personal interest is only *part* of one's action. The egoist takes this *part* for the whole, disregarding the social conditions that allow it to take place. Combining the cognitivist analysis of the the subjective maxim with a universalization procedure, does not produce the one true normative principle, but it does produce a powerful tool for refocusing evaluative perception allowing us to recognize when we fail to treat humanity with the dignity it deserves.

---

[238] Appiah makes it clear that the expansion of popular press in the form of print journalism played a big role in putting the practice of dueling under the public microscope and contributed to the broader process of democratization.

5.6.   Conclusion

In this chapter, it has been argued that the socialized perfectionist version of the self-conception

strategy can succeed with the aid of a theory of evaluative perception that prizes human freedom.

Attempts to prove the sociality thesis failed, but the general picture remains somewhat

compelling. The constitutivist alternative to perfectionism faces a serious obstacle in the form of

the constitutivist's dilemma. And conceptual resources like abstraction and FUL are available to

aid in focusing evaluative perception on salient features of ethical life.

CONCLUSION

Is Dueling Morally Permissible?

Probably not—but some questions are interesting if only for the fact that they are worth asking.

I'd like to conclude by revisiting Appiah's discussion of dueling in *The Honor Code,* in order to

illustrate how the view I've tried to develop in the second half of this dissertation may help to

deliver a verdict the question of dueling.[239]

The view I've been working toward is broadly perfectionist. It concerns an agent's

attempt to live up to an ideal. It is also a version of the self-conception strategy; the fundamental

question concerns the extent to which the agent approximates their idea. And, while the ideal is

the agent's *own* ideal, it is grounded in a socialized understanding of the self. One's idea of the

kind of person one wants to be is demarcated by a system of social relations; various objective

facts about one's relations to others determine the ways in which one can realistically conceive

of oneself. Instead of just seeing oneself as good rational agent, one tends to see oneself as a

good mother, dentist, or shrimp-boat captain. This *socialized* and *perfectionist* version of the *self-*

conception strategy is then combined with a theory of conscience that emphasizes a feeling of

awe that accompanies the contemplation of autonomy. A central question for every agent

concerns whether they are being sufficiently respectful to that awe-inspiring freedom that they

encounter within themselves and others.

To determine whether dueling is morally permissible, the crucial question is this: can a

participant in a duel conceive of himself (or herself) as an agent respecting the autonomy of

---

[239] See Chapter 5, Section 4.1.2

others according to the modes of expression available to them in the culture in which they operate? One problem has to be dismissed right away. Agents might conceive of themselves as expressing respect because they are delusional or oblivious to how others perceive their actions. This is where sociality and education come in. What our actions express is not entirely up to us, and part of moral education is learning what does and does not express respect.[240] An agent participating in a duel already exists within a network of social relations and social practices that already have meaning. The dueler cannot simply imagine that this or that duel is an ideal case; the meaning of the act must be understood in its social context.

So, the question becomes something like this: is the practice of dueling consistent with respect for autonomy and rational agency? We will see that the answer is a qualified "no." While there is no logical inconsistency in the notion of an ideal social context, in which dueling may be consistent with respect for autonomy. But this is not the England that Appiah is discussing.

Appiah maintains that honor is a "call upon us made by reason." When two people inhabit the same honor world, they are both governed by a set of standards reflected in that world's honor code.[241] It's not terribly difficult to understand the honor code that makes dueling seem like something noble. When we consider the priority that the present view places on respect, it seems fitting that the victim of a severe insult should be granted some form of apology. If apology is refused, it may even seem fitting that such an apology be demanded. Furthermore, there's a certain honor in standing up for oneself. While have seen that Kant associates *humility* with moral feeling, he explicitly rejects "false humility (servility)" as

---

[240] See Chapter 3, Section 4; Chapter 5, Section 5.

[241] Appiah, 2010: 183.

"directly contrary to [one's] character as a moral being" (**MdS** 6:420). From this perspective, avenging one's honor by dueling may seem justifiable. And, indeed*,* we find Kant somewhat conflicted by the issue. He associates it with murder, but he allows that it may sometimes be necessary to defend a soldier's honor. And, in such cases, Kant resists the idea that soldiers who duel should be punished (**MdS**, 6:336-37).

Rousseau opposed dueling but adopted a somewhat measured tone. He seems to have thought that banning the duel would just make things worse. The approach, instead, should be to attempt to redirect public opinion.[242] Appiah's account account in his history of moral revolutions bares this strategy out.[243] The moral arguments against dueling had been in place for a long time before the practice ended. Dueling was illegal in England long before the Duke of Wellington engaged in it. But this didn't stop even members of the political elite from dueling to defend their honor.[244] As long as dueling was an effective way to vindicate personal honor, the practice continued. What changed was not law or morality but notions of personal honor.

Appiah makes a crucial point concerning the structure of the duel: while dueling presupposed the equality of the participants, not everyone was able to participate. A gentleman would defend his honor in relation to another gentleman.[245] You take a risk in dueling. Sometimes you die. Sometimes you lose your job. Lord Canning, for instance, had to resign from

---

[242] See LaVeque Manty, 2006: 722-24 for a good discussion of Rousseau's remarks on dueling in "Letter to M. d'Alembert on the Theatre."

[243] Appiah, 2010: 1-53.

[244] "In taking the measure of Wellington's honor world and its norms we should notice that, of the ten men who preceded Wellington as prime minister—Lord Shelburne, William Pitt the Younger, and Canning—fought duels, as did Charles Fox and the Earl of Bath, each of whom was almost prime minster; and Peel, who eventually followed the Duke to the premiership, had shown himself willing to accept challenges" (Appiah, 2010: 20).

[245] It's worth noting that, when Rousseau and Kant talk about dueling, they tend to focus on soldiers. In the case of soldiers, it may be that the everyday proximity to death and the issue of formal rank complicate the issue a bit.

his position as foreign secretary. Hamilton died; Burr got to keep his job as veep.[246] This lax

regulation of the law reflects a certain agreement among the gentlemanly class, despite the

formal acknowledgement that dueling was a thing to be avoided. In risking life and status in

defense of honor, the participants demonstrated an absolute commitment to their deepest

values.[247] To participate in a duel, therefore, is to elevate oneself above earthly concerns. One

adopts a position in a kind of transcendent moral community, defined by a shared code of honor.

One's status as a respectable gentleman means more than life itself. And in challenging another

person to a duel, one necessarily acknowledges the other as (at least on some level) as an equal.

Their insults matter because they also belongs to this transcendent moral community. It would

have been absurd to accept a challenge from a person who was not also regarded as a gentleman.

Ultimately, it was the collapse of this moral community that lead to the end of dueling.

Appiah quotes Francis Bacon's more or less accurate prediction of the decline of the duel. Bacon

thought that "men of birth and quality" would stop dueling once the practice "came so low" as to

involve "bakers and base mechanical persons."[248] The increased democratization of English

social life in the early 19th century lead to the rise of a managerial class of businessmen and

bureaucrats. This new middle class brought with it a new culture of honor that eventually came

to dominate. So, although dueling involves mutual recognition among the participants, its

historical existence depended on the exclusion of outsiders who were understood as incapable of

participating in the culture of honor. When the middle class started doing it, the duel began to

---

[246] Appiah, 2010: 20-21.

[247] There is an obvious connection to be made with Hegel's discussion of the life and death struggle. "The relation of both self-consciousness is thus determined in such a way that it is through a life or death struggle that each *proves its worth* to itself, and that both *prove their worth* to each other" (**PhG**, p. 187-88)

[248] Quoted in Appiah, 2010: 46.

look crude. This may have had something to do with the dry technicality of bourgeois values.

The Duke of Wellington, as prime minister, may have looked comical in his duel because he was

playing for two audiences. Those concerned with his honor as a gentlemen may have admired his

concern for honor while being dismayed by the lowly state of modern politics. Those more

concerned with his technocratic role as leader of parliament may have wondered what this

pompous display had to do with getting laws passed. What's clear is that the values of the middle

class could no longer be easily dismissed, and in this new context, the duel was no longer an

effective means o asserting one's status as a respectable gentleman.

Dueling is an interesting case to consider from the point of view being advocated here. It

is far from obvious that the duel involves any failure of respect *between* the two participants. To

be sure, a given duel may be motivated by an initial failure of respect, but to engage in the duel

involves the assumption that the other is a recognized member of one's honor world, that the

other's opinion is worthy of consideration. It may be argued that dueling is wrong because it

dehumanizes the other, by reducing the other to a single transgression and taking their life

therefore. There may be a character flaw in the person who's quick to instigate a duel, but duels

may be the result of a long process, where the offender is given ample opportunity to apologize.

It seems reasonable to assume that if a duel takes place, both parties have decided to take their

stand on the issue. If this is so, then the duel—even if it results in death—may be consistent with

respect. Nevertheless, the practice of the duel involves a failure of respect for those who are

systematically *excluded* from the practice as if they are incapable of the same kind of moral

status. Excluding non-gentlemen from the community of respected moral agents is inconsistent

with respecting their humanity. So, from this we can conclude that the historical practice of

dueling was immoral because it reflected disrespect for moral agents who were excluded from the practice. And, if such exclusion is a necessary feature of the practice, then it follows that dueling is morally impermissible in general.

It's possible for an honor code to be defective at a general level. This is the case with an institution, like slavery, that fails to respect the inherent dignity and worth of humanity. It's also possible for an honor code to create specific conditions under which a given action shows up as a failure of respect. This is the case, at least, for dueling as it was historically practiced. Dueling was *objectively* wrong because it reflected a lack of respect for the humanity—if not that embodied in the life of the other person, then in those who were excluded from the moral community. Whether an inclusive system of dueling is possible is an empirical question that I will not attempt to answer here. Instead, I would argue that our our understanding of honor is so different from that of the early 19th century that such a system is just not a live option for us today. We have for far too long taken on that technocratic spirit that killed the duel in the first place. When someone cuts you off in traffic, they show you no regard, but you respond in kind—perhaps, with no more than a finger—and forget about it by the time you get to where you're want to go. For most of us, gone are the days in which an insult is insulting. F*or us*, taking a life for such a triviality would be dehumanizing. It would be to take the other person's insult to me for the whole of their being.[249] At a general level, dueling is wrong for us for the same reason it was wrong in early 19th century England, because it reflects a failure of respect. Today it's a failure to respect the dead; back then it was a failure to respect those who were prohibited from

---

[249] See the discussion of abstraction in Chapter 5, Section 5.1

participating. It's an open question whether there could ever be a permissible form of dueling, but the view developed here goes a long way in telling us what to look for.

So, the verdict on dueling is ambiguous. Under certain conditions, it *might* be possible for the practice to exhibit respect for rational agency. Under such conditions, dueling would be morally permissible, but we can be pretty sure that those are not the sorts of conditions of anyone reading this in 2019. The value of the view in question is that it directs our attention to what matters—the inner worth of our fellow human beings. We can understand, with a certain sympathy, why the nobility engaged in the practice, and we can also see how the duel became a joke. This sort of deep respect for cultural difference is not grounds for cultural relativism; it is evidence of a keen and broad-minded capacity for evaluative perception. We can see why it mattered for the nobility, and we can see why the duel collapsed under the weight of its haughty elitism. The class of people that it's exclusivity depended upon were made visible.

The Perfectionist Tradition and 18th and 19th Century Theories of Conscience

> "…the validity of duty is an other-worldly beyond of consciousness, [yet] this validity only comes about in consciousness…" (**PdG**, 632).

As rational beings, our nature is a matter of concern for us. And this concern extends beyond the biological organism that is the proper object of the natural sciences. In this dissertation, we have seen that by grounding moral philosophy in human agency and linking agency to sociality, a conception of what it is to be human begins to emerge. And to this conception belongs a characteristic form of evaluative perception: when one considers human freedom rightly, the appropriate attitude is one of reverence. Kant's discussion of freedom in the *Third Antinomy* already indicates an essential limit in the empirical investigation of causal relations in nature.

The idea of freedom furnishes the most indispensable element of a complex network of (mostly social) structures of meaning that give significance to a life. With this idea at the center, one develops a workable sense of self that serves as the basis of one's idea of the kind of person they (and feel) they ought to be. Respect for autonomy, combined with a critical understanding of the social structures and systems of cooperation that make our actions possible, provides guidepost for judging the rightness or wrongness of individual behavior.

Rousseau's egalitarian perfectionism turns ancient stoicism on its head, recasting the ideal of the sage in terms of the simple peasant virtues. To live well requires the ability to look upon one's life with the knowledge that one was good. This naturally leads to the broadly Kantian self-conception strategy, since being able to conceive of one's life is a necessary precondition for conceive of it as good. But it is a mistake to think that the logico-practical conditions consistent self-conception are more stringent and fundamental than those that shape the contours of our actual lives. Being able to regard oneself as a free rational agent may be a necessary condition for being able to regard oneself as a father. It doesn't follow that being a free rational agent is more central to one's self-conception. We can say, with Korsgaard, that immorality can make you "no longer be able to think of yourself under the description under which you value yourself and find your life to be worth living."[250] But logico-practical consistency in willing is not the basis for this ability.

I conclude with two examples that help illustrate the point: When Kant was ordered to write a letter to Frederick Wilhelm II promising to abstain from the public discourse on religion, he famously signed, "as your majesty's most loyal subject." After the Frederick Wilhelm's death,

---

[250] See section 5.3.2.

Kant maintained that he was no longer bound by his promise because the carefully worded signature no longer held.[251] I don't know what was in Kant's mind when he signed the letter, and I don't know if this legalistic waffling makes his action consistent with the universal law formulation of the categorical imperative. But if we suppose that morality is about being able to think of oneself under one's most valued description, then it's worth asking. Which Kant is easier to understand and more deserving of respect? The Kant who tinkers with legalese and abstract principles in order to 'keep' a promise he clearly believes he shouldn't have made in the first place? Or the one who stands on principle in a different way (as a more Jacobin Kant might have done), by openly breaking his promise and placing a more concrete principle of liberty above the abstract moral law? To become intelligible, freedom needs a hill worth dying on, and though many have died for the right to believe freely, few would stand firm on mere consistency.

The second example is more recent. It concerns the ex-white-supremacist, Derrick Black. For the first twenty years of his life, Derrick was groomed by his father Don Black—founder of the racist website, *Stormfront*—to be a leader int the white-supremacist movement. At eighteen, Derrick moved away to college and made a series of connections with other young adults. Through these relationships Derrick was confronted with a certain kind of inconsistency. His self-conception was grounded on the demonization of those who were racially and culturally different from him. This demonization allowed him to view himself as a kind of hero, protecting his race from so-called "white genocide." His new friends found this view of things grotesque, and their undeniable decency made their demonization unsustainable.[252] In the end, Derrick was

---

[251] For a nice discussion of this controversy, see Hunter, 2005.

[252] One is struck by the example of the Jewish students on campus, who made a habit of inviting Derrick to their Sabbath dinner. See Saslow, 2018.

no longer able to think of himself under this valued description—not because of any logical

inconsistency in willing but because that description was founded on a network of untruths. It

was simply incompatible with the reality with which we was now faced. I don't know what was

in Derrick's mind when he abandoned white-nationalism, but one must assume that such a

profound reimagining of what makes life worth living goes far beyond the mere logic of moral

principle. Charles Mill's interpretation of Kant shows how easy it is to make racism *logically*

consistent.[253] But when morality is answerable to one's *actual* self-conception, perfectionist

considerations become more salient and tribalism is harder to sustain.

In the *First Critique*, Kant illustrates the distinction between *idea* and *ideal* with the

example of the stoic sage: "virtue, and, with it, human wisdom in its absolute purity, are ideas.

But the stoic sage is an ideal—that is, a person existing in mere thought but that fully coincides

with the idea of wisdom."[254] Recall the basic idea behind Kant's moral postulates. We know

enough about morality to know that happiness and virtue ought to go hand-in-hand. Whatever

obligates us must be actually achievable (not just logically possible). The existence of god,

eternal life, and freedom are necessary for the achievement of the highest good. Therefore, they

must be actual. The *idea* of the highest good contains these elements as part of its concept, but

the *ideal* is the concept imagined in its fullness—something like a kingdom of God. In *The Spirit*

of Christianity and its Fate, Hegel takes up this distinction and incorporates it into a conception

of religious practice.

> Religious practice is the most holy, the most beautiful of all things; it is our
> endeavor to unify the discords necessitated by our development and our attempt

---

[253] Mills, 2005. See introduction for more on Kant's racism.

[254] **KrV**, A569/B597.

> to exhibit the unification in the *ideal* as fully *existent*, as no longer opposed to reality, and thus to express and confirm it in a deed.[255]

Notice that on Hegel's view, contradictions inherent in existence don't stop us from striving after the ideal. Indeed, we find at the end of the essay that "the fate" of Christianity is that it can *never* fully reconcile existence with the ideal. He contrasts the modern condition with that of the apostles, whose "cognition is more like he vague hovering between reality and spirit." The figure of Jesus, as the union of man and deity, is the manifestation of the great reconciliation of existence and the ideal. Instead it is the fate of Christianity to restlessly oscillate between the two, unable to find peace.[256]

We saw that Rousseau's egalitarian perfectionism begins and ends in tension with itself. The idea of freedom as a kind of rupture within oneself leads to a conception of achievement that involves recognizing that the greatest among is us no better than the least among us. The ideal of sagacity may be beyond our reach, but as the model, it has critical force. Ways of life reflect ways of responding to the infinite complexity at the heart of the human experience; the infinite diversity of such responses is both inspiring and to be expected. As Appiah reminds us:

> Our ancestors have been human for a long time. If a normal baby girl born forty-thousand years ago were kidnapped by a time-traveler and raised by a normal family in New York, she would be ready for college in eighteen years. She would learn English (along with—who knows?—Spanish or Chinese), understand trigonometry, follow baseball and pop music; she would probably want a pierced tongue and a couple of tattoos. She would be unrecognizably different from the brothers and sisters she left behind.[257]

---

[255] **ETW**, 206.

[256] **ETW**, 300-301.

[257] Appiah, 2006: xi.

It's an old mistake, exaggerating the threat of Darwinism to morality. In its simplest form, the great moral lesson of the Darwinian revolution goes like this: unlike some myths about a divine creator, the real origin of human nature probably didn't make us good.

The philosophers discussed here did not rest their views on the assumption that our humanity makes us *good.* Our humanity makes it possible for us to be *evil*. Human freedom creates the possibility of a radical disharmony within the self and the lifelong project of coming to terms with that condition. Indeed, it is humanism and not relativism that results when we accept Lessing's famous formulation:

> If God were to hold all truth concealed in his right hand, and his left only the steady and infinite striving after truth, albeit with the proviso that I would always and forever err in the process, and to offer me that choice, I would with all humility take the left hand and say: Yea, the pure truth is but for you alone![258]

---

[258] Lessing, 1897: 23-24.

BIBLIOGRAPHY

Abernathy, Charles and Robert Hamm. *Surgical Intuition: What it is and how to get it.* Philadelphia: Hanley and Belfus, 1994.

Allen, W. Stannard. *Living English Structure: Practice Book for Foreign Students.* Fourth Edition, London: Longman, 1959.

Allais, Lucy. "Kant's Racism." *Philosophical Papers,* vol. 45, nos. 1-2, 2016: 1-36.

Appiah, Kwame Anthony. *Cosmopolitanism: Ethics in a World of Strangers.* New York: W.W. Norton and Company, 2006.

---. *The Honor Code: How Moral Revolutions Happen.* New York: W.W. Norton and Company, 2010.

Aristotle. *Nicomachean Ethics.* Translated and Edited by Terence Irwin Second Edition. Indianapolis: Hackett Publishing, 1999.

Audi, Robert. *Moral Perception.* Princeton, NJ: Princeton University Press, 2013.

—. "Intuition and Its Place in Ethics," *Journal of the American Philosophical Association.* 1, no. 1 (2015): 57-77.

—. *The Good and the Right: A Theory of Intuition and Intrinsic Value.* Princeton: Princeton UP, 2005.

Ayer, A.J. *Language, Truth and Logic.* Dover, 1952.

Babbitt, Irving. *Rousseau and Romanticism.* New York: Houghton Mifflin, 1919.

Beiser, Frederick. *The Fate of Reason: German Philosophy from Kant to Fichte.* Cambridge: Harvard University Press, 1987.

—."Hegel: A Non-Metaphysician? A Polemic!" *The Bulletin of the Hegel Society of Great Britain.* Vol 32. 1995: 1-13.

—."Dark Days: Anglophone Scholarship Since the 1960's" in *German Idealism: Contemporary Perspectives*, ed. Aspen Hammer. New York: Routledge 2007: 70-90.

Bernasconi, Robert. "Who Invented the Concept of Race? Kant's Role in the Enlightenment Construction of Race," in Robert Bernasconi ed. *Race,* Oxford: Blackwell, 2001: 11-36.

—."Will the Real Kant Please Stand Up: The Challenge of Enlightenment Racism to the Study of the History of Philosophy," *Radical Philosophy*, vol. 117, jan./feb. 2003. 13-22.

Blackburn, Simon. *Ruling Passions: A Theory of Practical Reasoning.* Oxford: Oxford University Press, 1998.

Bloom, Paul. *How Pleasure Works: The New Science of Why we Like What we Like.* New York: W.W. Norton and Company, 2010.

—.*Against Empathy: The Case for Rational Compassion.* New York: Harper Collins, 2016.

Bogardus, Tomas. "Only All Naturalists Should Worry about Only One Evolutionary Debunking Argument," *Ethics* vol. 126, 2016: 636-61.

Bosker, Bianca. *Cork Dork: A Wine Fueled Journey into the Art of Sommeliers and the Science of Taste.* London: Allen & Unwin, 2017.

Boyd, Robert and Peter Richerson. *Not by Genes Alone: How Culture Transformed Human Evolution.* Chicago: University of Chicago Press, 2005.

Boyd, Richard N. "How to be a Moral Realist" in Sayre-McCord (ed) *Essays on Moral Realism.* Ithaca, New York: Cornel University Press, 1988: 181-229.

Cassirer, Ernst. *Rousseau Kant Goethe: Two Essays.* Translated by James Gutmann, Paul Oskar Kristeller, and John Herman Randall Jr. Princeton, NJ: Princeton University Press, 1945.

—. *The Question of Jean Jacques Rousseau.* 2nd ed. Edited and translated by Peter Gay. New Haven, CT: Yale University Press, 1989.

Churchland, Patricia. *Braintrust: What Neuroscience Tells Us about Morality*. Princeton: Princeton University Press, 2018.

Clark-Doan, Justin. "Moral Epistemology: The Mathematics Analogy," *Nous* 48, no. 2, 2014: 238-55.

—."Morality and Mathematics: The Evolutionary Challenge," *Ethics* 122, no. 2, 2012: 313-40.

Copp, David. "Moral Naturalism and Three Grades of Normativity," in *Normativity and Naturalism.* Ed Paul Shraber. Frankfurt: Ontos Verlag, 2004: 1-45.

Correll, J., Park, B., Judd, C.M., and Wittenbrink, B. "The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals," *Journal of Personality and Social Psychology* 83 no. 6 (2002): 1314-29.

Correll J., Wittenbrink, B., Crawford, M.T. and Sadler, M.S. "Stereotypic Vision: How Stereotypes Disambiguate Visual Stimuli," *Journal of Personality and Social Psychology*, 108, no. 2 (2015): 219-33.

Cullison, Andrew. "Moral Perception," *European Journal of Philosophy* 18, no. 2, 2009: 159-175.

Deleuze, Gilles. "Michel Tournier and World without others," translated by Graham Burchell, *Economy and Society.* vol. 13, no. 1, 52-71.

DiGiorgio-Miller, Janet. "Sibling Incest: Treatment of the Family and the Offender," *Child Welfare*, vol. 77, no. 3, 1998: 335-46.

Douglass, Frederick. *Narrative of the Life of Frederick Douglass, An American Slave Written by Himself.* Eds. John Blassingame, John McKivigan, and Peter Hinks. New Haven: Yale University Press, 2016.

Epstein, Seymour. "Intuition: What it is, What it does, and how it does it," *Psychological Inquiry,* vol. 21, no. 4, 2010: 295-312.

Felser, Daniel and Carlos Navarrete. "Third-Party Attitudes toward Sibling Incest: Evidence for Westermark Hypothesis," *Evolution and Human Behavior,* vol. 25, 2004, 277-94.

Fitzpatrick, William. "Valuing Nature Non-Instrumentally," *The Journal of Value Inquiry.* vol. 38, 2004, 315-32.

—."Debunking Evolutionary Debunking of Ethical Realism" in *Philosophical Studies.* vol. 172, 2015, 883-904.

Forbes, Robert. "Slavery and the Evangelical Enlightenment," in *Religion and the Antebellum Debate over Slavery*, Eds. John McKivigan and Mitchell Snay. Athens: University of Georgia Press, 1998: 68-108.

Forster, Michael. *Kant and Skepticism*. Princeton: Princeton University Press: 2008.

Foucault, Michel. *The History of Sexuality, Volume One: Introduction*. trans. Robert Hurley, New York: Pantheon Books, 1978.

Freud, Sigmund. *The Future of an Illusion.* Translated by Peter Gay. New York: W.W. Norton & Company, 1989, [1927].

—.*Civilization and its Discontents*. Translated by Peter Gay. New York: W.W. Norton & Company, 1989, [1930].

Gibbard, Allan. "Morality as Consistency in Living: Korsgaard's Kantian Lectures," *Ethics,* 110, October, 1999: 140-64.

Giovanni, George di. "The First Twenty Years of Critique: The Spinoza Connection" in Paul Guyer (ed) *The Cambridge Companion to Kant.* Ed. Paul Guyer. Cambridge: Cambridge University Press, 1992.

Greco, John. *Putting Skeptics in Their Place: The Nature of Skeptical Arguments and Their Role in Philosophical Inquiry.* Cambridge: Cambridge University Press, 2000.

Gregory, R.L. "Distortion of Visual Space as Inappropriate Constancy Scaling," *Nature,* 1963.

Goldstein, Robin, et. al. "Do More Expensive Wines Taste Better? Evidence from a Large Sample of Blind Tastings," *AAWE Working Paper,* no. 16, April 2008.

Gould, Stephen Jay. "Sociobiology: The Art of Storytelling," *New Scientist,* 16 November 1978.

Hacking, Ian. "The Looping Effects of Human Kinds," in *Causal Cognition: an Interdisciplinary Approach,* Eds. D.T. Wasserman and R.T. Wachbroit. Oxford: Oxford University Press, 1995: 351-83.

Haidt, Jonathan. "The Emotional Dog and its Rational Tail," *Psychological Review,* vol.108, no. 4, 2001 814-34.

—. *The Happiness Hypothesis: Putting Ancient Wisdom and Philosophy to the Test of Modern Science.* London: Random House, 2006.

—. *The Righteous Mind: Why Good People are Divided by Politics and Religion.* New York: Random House, 2012.

—. "The Emotional Dog Learns New Tricks," *Psychological Review,* Vol. 110, No. 1, 2003: 197-98

Haidt, Jonathan and Craig Joseph. "How Moral Foundations Theory Succeeded in Building a House on Sand: A Response to Suhler and Churchland," *Journal of Cognitive Neuroscience,* vol. 23, no. 9, 2011: 2117-22.

Hanisch, Christoph. "Constitutivism and Inescapability: A Diagnosis," *Philosophia,* vol 44, 2016: 1145-64.

Hare, John. *The Moral Gap: Kantian Ethics, Human Limits, and God's Assistance.* Oxford: Clarendon Press, 1997.

—. "Kant on the Rational Instability of Atheism," in *Kant and the New Philosophy of Religion* Eds. Chris Firestone and Stephen Palmquist. Bloomington: Indiana University Press, 2006: 62-78.

Harman, Gilbert. *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press, 1977.

Henrich, Joseph. *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating Our Species and Making Us Smarter*. Princeton: Princeton University Press, 2016.

Hobbes, Thomas. *The Leviathan: with Selected Variants from the Latin Edition of 1668.* ed. Edwin Curley. Indianapolis: Hackett, 1994.

Hunter, Ian. "Kant's *Religion* and Prussian Religious Policy," in *Modern Intellectual History,* vol. 2, no. 1, 2005: 1-27.

Jaeger, Gary. "The Wisdom of Hindsight and the Limits of Humean Constructivism," *Philosophical Explorations*, vol. 18, no. 1, 2015: 45-57.

Jaeggie, Rahel. *Alienation*. Trans. Frederick Neuhauser and Allen Smith. Ed. Frederick Neuhauser. New York: Columbia University Press, 2016.

—. *Kritik von Lebensformen.* Berlin: Suhrkamp Verlag, 2014.

Joyce, Richard. *The Evolution of Morality*. Cambridge: MIT Press, 2006.

—-. *The Myth of Morality.* Cambridge: Cambridge University Press, 2001.

Juvenal. *The Satires*. Translated by Niall Rudd. Oxford: Oxford University Press, 1991.

Kahneman, Daniel. *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux, 2011.

Kahneman, Daniel and Gary Klein. "Conditions for Intuitive Expertise: A Failure to Disagree," *American Psychologist,* vol. 64, no. 6, 2009: 515-26.

Kain, Patrick. "Realism and Anti-Realism in Kant's Second Critique." *Philosophy Compass*, vol. 1 no. 5, 2006: 449-65.

---"Self-Legislation in Kant's Moral Philosophy" *Archiv für Geschichte der Philosophie*, vol. 86, bd. S, 2004: 257-306.

Kass, Leon and James Wilson. *The Ethics of Human Cloning.* Washington DC: The AEI Press, 1998.

Kaufmann, Walter. *Hegel: A Reinterpretation, Texts, and Commentary*. New York: Doubleday, 1965.

Kelly, Dan. *Yuck! The Nature and Significance of Disgust.* Cambridge: MIT Press, 2011.

---*Projectivism Psychologized: The Philosophy and Psychology of Disgust*. Thesis—(Phd), Rutgers University, 2007.

Kelly, Dan, E. Machery, and R. Mallon. "Race and Race and Racial Cognition" in Doris et. al. Eds. *The Moral Psychology Handbook.* New York: Oxford University Press, 2010a.

Kelly, Dan. L. Faucher and E. Machery. "Getting Rid of Racism: Addressing Three Proposals in Light of the Psychological Evidence." *Journal of Social Philosophy,* 41, no. 3, 2010b, 292-322.

Keltner, Dacher and Jonathan Haidt. "Approaching Awe: A Moral, Spiritual, and Aesthetic Emotion," in *Cognition and Emotion*, vol. 17, no. 2, 2003:

Klein, Gary. *Sources of Power: How People Make Decisions.* Cambridge: MIT Press, 1999.

Korsgaard, Christine. *The Sources of Normativity.* Cambridge: Cambridge University Press, 1996a.

—. "Kant's analysis of obligation: The argument of *Groundwork I*" in *Creating the Kingdom of Ends,* Cambridge: Cambridge University Press, 1996b: 43- 76.

—. "Kant's Formula of Universal Law," in *Creating the Kingdom of Ends,* Cambridge: Cambridge University Press, 1996b: 77-105.

—. "Introduction" in *Groundwork for the Metaphysics of Morals*, revised edition. eds. Mary Gregor and Jens Timmerman. Cambridge: Cambridge University Press, 1998: ix-xxxi.

—. "Self-Constitution in the Ethics of Plato and Kant," in *The Journal of Ethics*, vol. 3, 1999: 1-29.

—. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press, 2009.

Laitinen, Arto. *Strong Evaluation Without Moral Sources: On Charles Taylor's Philosophical Anthropology and Ethics*. Berlin: de Gruyter, 2008.

—"Hegelian Constructivism in Ethical Theory?" In *I That is We, We that is I. Perspectives on Contemporary Hegel.* Eds. Italo Testa and Luigi Ruggiu. Boston: Brill Publishers, 2016: 127-46.

LaVaque-Manty, Mika. "Dueling for Equality: Masculine Honor and the Modern Politics of Dignity." in *Political Theory,* vol. 34, no. 6, 2006: 715-40

Lawson, Rebecca. "The Science of Cycology: Failures to Understand How Everyday Objects Work," *Memory and Cognition,* vol. 38, no. 8, 2006: 1667-75.

Levine, Alan. *The Politics of Autonomy: A Kantian Reading of Rousseau's Social Contract*. Amherst: University of Massachusetts Press, 1976.

Lessing, Gotthold Ephraim. "Eine Duplik," in *Sämtliche Schriften,* Vol. 13, Ed. Karl Lachman. Leipzig: G.J. Göschen, 1897: 21-93.

Lieberman, Debra, John Tooby, and Leda Cosmides. "Does Morality Have a Biological Basis? An Empirical Test of Factors Governing Moral Sentiments Relating to Incest," *Biological Sciences*, vol. 270, no. 1517, 2003: 819-26.

Louden, Robert. "Not a Slow Reform, but a Swift Revolution: Kant and Basedow on the Need to Transform Education," in *Kant and Education: Interpretations and Commentary.* Eds. Klas Roth and Chris Surprenant. New York: Routledge, 2011: 39-54.

Machery, Edouard and Ron Mallon. "The Evolution of Morality," in *The Moral Psychology Handbook* Ed. John Doris. Oxford: Oxford University Press, 2009: 3-46.

Mackie. J.L. *Ethics: Inventing Right and Wrong.* London: Penguin Books, 1977

Mazzucat0, Mariana. *The Value of Everything: Making and Taking in the Global Economy.* London: Allen Lane, 2018.

McGrath, Sarah. "Moral Knowledge by Perception," *Philosophical Perspectives*, vol. 18, 2004: 2009-28.

Meacham, John. *American Gospel: God, The Founding Fathers, and the Making of a Nation.* New York: Random House, 2006.

Mill, Jon Stuart. *Utilitarianism, and the 1868 Speech on Capital Punishment.* 2nd Edition. Indianapolis: Hackett Publishing, 2002.

Mills, Charles. "Kant's *Untermenschen*" in Andrew Valls, *Race and Racism in Modern Philosophy*, New York: Cornell University Press, 2005: 169-93.

Montaigne, Michel de. *The Complete Essays of Michel de Montaigne.* Trans. Donald Frame. Stanford: Stanford University Press, 1958.

Montesquieu, Charles Baron de. *Spirit of the Laws*. Eds. Anne Cohler, Basia Caroline Miller, and Harold Samuel Stone. Cambridge: Cambridge University Press, 1989.

Moore, G.E. *Principia Ethica.* Cambridge: Cambridge University Press, 1903.

Mundy, Matthew. "Testing day: The effects of processing bias induced by Navon stimuli on the strength of the Mueller-Lyer illusion" *Advances in Cognitive Psychology*, 2014, vol. 10(1): 9-14.

Nietzsche, Friedrich. *Werke in Sechs Bänden*. 5th Edition. Ed. Karl Schlechta. Munich: Carl Hanser Verlag, 1980.

O'Neill, Onora. "Vindicating Reason," *The Cambridge Companion to Kant.* Edited by Paul Guyer. Cambridge: Cambridge University Press, 1992.

—. "Kant and the Social Contract Tradition" in *Kant Actual: Homage a Pierre LaBerge.* Eds. Francois Duchesneau, Guy LaFrance, and Claude Piche. Paris: Librarie Philosophique J. Vrin, 2000.

—. "Constructivism in Rawls and Kant," *The Cambridge Companion to Rawls*. Cambridge: Cambridge University Press, 2006.

Payne, B. K. "Prejudice and Perception: The Role of Automatic and Controlled Processes in Misperceiving a Weapon," *Journal of Personality and Social Psychology* 81, no. 2. (2001): 181-92.

Pinkard, Terry. *German Philosophy 1760-1860: The Legacy of Idealism.* Cambridge: Cambridge University Press, 2002.

Pippen, Robert. *Modernity as a Philosophical Problem: On the Dissatisfactions of European High Culture.* Oxford: Blackwell, 1991.

—. *Hegel's Practical Philosophy: Rational Agency as Ethical Life.* Cambridge: Cambridge University Press, 2008.

Pinker, Steven. *How the Mind Works.* New York: W.W. Norton and Co., 1997.

---*Better Angels of Our Nature: Why Violence has Declined.* New York: Penguin Books, 2011.

---*Enlightenment Now: The Case for Reason, Science, Humanism, and Progress.* New York: Random House, 2018.

Plantinga, Alvin. *Warrant and Proper Function.* Oxford: Oxford UP, 1993.

Plassman, Hilke et. al. "Marketing Actions can Modulate Neural Representations of Experienced Pleasantness," *PNAS,* vol. 105 no. 3, 2008: 1050-54.

Plato. *Republic*. Translated by C.M.A. Grube. Indianapolis: Hackett Publishing Company, 1992.

Prinz, Jesse. "The Emotional Basis of Moral Judgment," *Philosophical Explorations*, vol. 9, no. 1, 2006, 29-43.

---*Beyond Human Nature: How Culture and Experience Shape our Lives.* New York: W.W. Norton, 2012.

---"Against Moral Nativism" in *Stitch: And His Critics.* Eds. Dominic Murphy and Michael Bishop. New York: Blackwell, 2009: 167-89.

Rae Micheal, "Naturalism and Moral Realism" in T.M. Crisp, M. Davidson, and D Vander Lann eds. *Knowledge and Reality,* (Dordrecht: Springer, 2006), 215-241.

Redding, Paul. *Analytic Philosophy and the Return of Hegelian Thought*, Cambridge: Cambridge University Press, 2007.

Richardson, Peter and Robert Boyd, *Not By Genes Alone: How Culture Transformed Human Evolution.* Chicago: University of Chicago Press, 2005.

Ross, Sir David. *The Right and the Good*. Oxford: Oxford University Press, 1930.

Rozenblit, Leonid and Frank Kiel. "The Misunderstood Limits of Folk Science: An Illusion of Explanatory Depth," *Cognitive Science,* vol. 26, 2002: 521-62.

Rozin, Paul. "The Process of Moralization." *Psychological Science*, vol. 10, no. 3, 1999: 218-21.

Rozin, Paul , Jonathan Haidt, Clark McCauley, Lance Dunlop, and Michelle Ashmore. "Individual Differences in Disgust Sensitivity: Comparisons and Evaluations of Paper-to-Pencil versus Behavioral Measures," *Journal of Research in Personality.* vol. 33, 1999: 330-51.

Russell, Bertrand. *The Problems of Philosophy*. London: Oxford University Press, 1959.

---*A History of Western Philosophy*. New York: Routledge, 2004.

Saslow, Eli. *Rising Out of Hatred: The Awakening of a Former White Nationalist*. New York: Doubleday, 2018.

Saunders, Benjamin, Julie Lipovsky and Rochelle Hanson. "Couple and Familial Characteristics of Father-Child Incest Families," *Journal of Family and Social Work*, vol. 1, no. 2 1995: 5-25.

Sayre-McCord, Geoffrey (Ed). *Essays on Moral Realism*. New York: Cornell UP, 1988.

---Sayre-McCord, Geoffrey. "Introduction: The Many Moral Realisms" in Sayre-McCord 1988.

Schafer-Landau, *Moral Realism: A Defense*. Oxford: Oxford University Press, 2003.

Schmidt, Marco F. H. and Michael Tomasello. "Young Children Enforce Social Norms," *Current Directions in Psychological Science*. vol. 21 no. 4, 2012.

Sellars, Wilfred. *Empiricism and the Philosophy of Mind*. Cambridge MA: Harvard University Press, 1997.

Simon, Herbert. "What is an 'Explanation' of Behavior?," *Psychological Science,* vol. 3 no. 3, 1992: 150-61.

Sloman, Steven and Philip Fernbach. *The Knowledge Illusion: Why we Never Think Alone,* New York: Riverhead Books, 2017.

Siegel, Susanna. *The Rationality of Perception.* Oxford: Oxford University Press, 2017.

Singer, Peter. "Famine Affluence and Morality" Morality," *Philosophy and Public Affairs* Vol. 1, No. 3, (Spring 1972), 229-43.

Street, Sharon. "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies,* vol. 127, 2006: 109-66

—. "Evolution and the Normativity of Epistemic Reasons," *Canadian Journal of Philosophy,* vol. 39, no. S1, 2009: 213-48.

—. "Mind Independence Without the Mystery: Why Quasi Realists Can't Have it Both Ways," in *Oxford Studies in Metaethics,* Vol. 6. Oxford: Oxford University Press, 2011: 1-32.

Streumer, Bart. *Unbelievable Errors: An Error Theory About Normative Judgments.* Oxford: Oxford University Press, 2017.

Summers, Jesse. "Rationalizing our Way to into Moral Progress," *Ethics, Theory, and Moral Practice*, vol. 20, 2017: 93-104.

Thaler, Richard and Cass Sunstein. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New York: Penguin Books, 2009.

Tidefors, Inga, Hans Arvidsson, Sara Ingevaldson, and Michael Larsson. "Sibling Incest: A Literature Review and a Clinical Study," *Journal of Sexual Aggression,* vol. 16, no. 3, 2010: 347-60.

Tooley, Michael. "Abortion and Infanticide," *Philosophy and Public Affairs,* vol. 2, no. 1, 1972: 37-65.

Watkins, Michael and Kelly Dean Jolley. "Pollyanna Realism: Moral Perception and Moral Properties." *Australasian Journal of Philosophy,* 80, no. 1 (2002): 75-85.

Welsh, Alexander. *What is Honor?: A Question of Moral Imperatives.* New Haven: Yale University Press, 2008.

Westphal, Kenneth. *Kant's Transcendental Proof of Realism*. Cambridge: Cambridge University Press, 2004.

—. "Kant's Anti-Cartesianism," *Dialogue: Canadian Journal of Philosophy/Revue canadienne de philosophie,* vol. 46, no. 4, 2007: 709-15.

—. "Natural Law, Social Contract and Moral Objectivity: Rousseau's Natural Law Constructivism," *Jurisprudence,* vol. 4.1, 2013: 48-75.

—. *How Hume and Kant Reconstruct Natural Law: Justifying Strict Objectivity* without *Debating Moral Realism.* Oxford: Oxford University Press, 2016.

Wood, Allen. *Hegel's Ethical Thought.* Cambridge: Cambridge University Press, 1990.

—. "Hegel and Marxism," in *The Cambridge Companion to Hegel,* edited by Frederick Beiser. Cambridge: Cambridge University Press, 1993: 414-44.

—. *Kantian Ethics.* Cambridge: Cambridge University Press, 2008.

Williams, Bernard. "On Internal and External Reasons," in *Moral Luck: Philosophical Papers, 1973-1980.* Cambridge: Cambridge University Press, 1981: 101-14.

Worling, James R. "Adolescent Sibling-Incest Offenders: Differences in Family and Individual Functioning When Compared to Adolescent Nonsibling Sex Offenders," *Child Abuse and Neglect,* vol. 19, no. 5, 1995: 633-43.

Wright, Robert. *Why Buddhism is True*. New York: Simon and Schuster, 2017.

Yeomans, Chris. *The Expansion of Autonomy.* Oxford: Oxford University Press, 2016.

—. "Identity as a Process of Self-Determination in Hegel's Logic," in *Identity and Difference.* Ed. Philip Grier. SUNY Press, 2007: 63-82.

Zöller, Günther. *Fichte's Transcendental Philosophy: The Original Duplicity of Intelligence and Will.* Cambridge: Cambridge University Press, 1998.