Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Bree Beal Date

Raskolnikov's Confession: A Dostoevskian Model of Moral Psychology

By

Bree Beal
Doctor of Philosophy

Graduate Institute of the Liberal Arts

_____
Kevin Corrigan
Advisor

_____
Philippe Rochat
Committee Member

_____
Cynthia Willett
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Raskolnikov's Confession: A Dostoevskian Model of Moral Psychology

By

Bree Beal
B.M., Grand Canyon University, 2004
M.A., Arizona State University, 2011

Advisor: Kevin Corrigan, Doctor of Philosophy

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in The Graduate Institute of the Liberal Arts
2018

Abstract

Raskolnikov's Confession: A Dostoevskian Model of Moral Psychology
By Bree Beal

In this work, I show that Dostoevsky's portrayal of the complex, or "polyphonic," dynamics of moral cognition issues a stark challenge to current theory in the field of moral psychology. While my thesis is informed by Dostoevsky's work taken broadly, I focus in on a fifteen-page passage in *Crime and Punishment*. In this passage, the murderer Raskolnikov confesses his crime to Sonya, a young woman who had been a close friend of one his victims. Raskolnikov offers seven distinct explanations for the double murder, each of which Sonya rejects. I argue that each of Raskolnikov's seven accounts is true but incomplete, and that the real moral-cognitive dynamics of Dostoevsky's account only emerge when all seven are taken together in all their inconsistency. I show that the confession—both the sundry explanations of the crime and the interpersonal dynamics of the conversation itself—expresses all the major features of an original model of moral psychology. Then I place this model in critical conversation with existing moral psychological research and theory. The major innovation of this model is that I explain the dynamics of moral cognition in terms of more basic ontological dynamics that give rise to morality: our relationships with valued people, places, creatures, objects, and ideas. I show that the moral domain is framed and reframed at each moment by our evolving ontologies. For instance, one assumes moral responsibilities simply by *being* a sister, daughter, soldier, student, citizen, and so on. One assumes certain moral responsibilities towards friends, pets, fellow humans, homeland, ideals, hopes and dreams, sacred objects, and so on. One's moral responsibilities fluctuate with memory, attention, mood, and situational features. I call the normative-ontological understanding that shapes moral judgment "existential framing," and I show that existential framing is the primary source of the polyphonic dynamics on display in Raskolnikov's confession and operative in everyday moral cognition.

Raskolnikov's Confession: A Dostoevskian Model of Moral Psychology

By

Bree Beal
B.M., Grand Canyon University, 2004
M.A., Arizona State University, 2011

Advisor: Kevin Corrigan, Doctor of Philosophy

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in The Graduate Institute of the Liberal Arts
2018

Acknowledgements

I want to thank my family: Mom and Dad, Shea, Rane, and Ty, for loving me and believing in me and being steadfast in supporting me. Carmen, for introducing me and my friends to Dostoevsky. Carmen and Nicki, for encouraging me to go back to school, and for helping me along the way. Nate, for talking Dostoevsky with me in those early days. Jason, for talking about all sorts of shit and helping me get into grad school. Justin, for the friendship and encouragement. Lily, for becoming my conscience. Carl, for being a special guy. Patrick, for almost single-handedly getting me into Emory and for being a great mentor and friend along the way. Kevin, for being the chillest and sweetest advisor. Cyndi and Philippe, for being the chillest and sweetest committee members. The Dilks lab, for all the good times. Elena and Peter and the IDS Writing Labs team, for your faith and support. The staff and faculty at the Institute for the Liberal Arts and the folks at the Fox Center, especially Lolitha, Martine, and Keith. Those who have influenced my thinking and writing simply by being beautiful thinkers and writers yourselves. Those who have inspired me with their accomplishments, their beautiful ways of living, their boldness, bravery, integrity, passion, and commitment. Those who have helped me survive multiple breakups, double-digit moves, and all the natural stresses of graduate student life, whether through logistical assistance or simply through being my friends. The city of Atlanta. All that is good or beautiful. I acknowledge my debt. I offer my gratitude.

Table of Contents

Chapter I: Raskolnikov's Confession

Chapter I: Contents

After her first passionate and tormenting sympathy for the unhappy man, the horrible idea of the murder struck her again. In the changed tone of his words she suddenly could hear the murderer. She looked at him in amazement. As yet she knew nothing of why, or how, or for what it had been. Now all these questions flared up at once in her consciousness. And again she did not believe it: "He, he a murderer? Is it really possible?"[1]

Introduction to the Dostoevskian Model

Early in Fyodor Dostoevsky's novel *Crime and Punishment,* Rodion Romanovich Raskolnikov receives a long letter from his mother, in which she tells of his sister Dunya's imminent decision to marry a wealthy older man, who—quite out of the blue—has recently proposed.[2] Pyotr Petrovich Luzhin, Raskolnikov understands from certain clues in the letter, is in all the bad senses a pig,[3] and Dunya could only consider a match with him out of desperation, in order to save her mother and brother from poverty. Distraught, Raskolnikov vows to stop the marriage from taking place.[4] But what can he—a law school dropout with no immediate prospects—do to stop it? The very next day, Raskolnikov murders and robs a wealthy pawnbroker, Alyona Ivanovna, also killing the pawnbroker's younger sister Lizaveta, who happens upon the scene of the crime.[5]

Raskolnikov faces external pressure from the clever detective Porfiry Petrovich, who strongly suspects him of the murder, but Raskolnikov is more distressed by his inability to resolve certain internal conflicts. Suffering from extended bouts of physical illness and feeling

---

[1] Fyodor Dostoevsky, *Crime and Punishment,* trans. Richard Pevear & Larissa Volokhonsky (New York: Vintage Classics, 2012), 412.

[2] Ibid.*,* 34-35.

[3] I use this word because it is currently our most vivid metaphor for the kind of capitalistic greed and male chauvinism that Luzhin embodies, but the comparison is terribly unjust to actual pigs.

[4] Dostoevsky, *Crime*, 40.

[5] Ibid., 76-79.

his sanity unraveling, he is impelled to take a decisive step, choosing either to kill himself or turn himself over to the police. Before he can make this decision, however, Raskolnikov is drawn to confess his crime to Sonya Marmeladov, a deeply religious and compassionate young woman of eighteen years, who has recently taken up prostitution in order to provide for her own impoverished family—that is, her drunkard father, abusive stepmother, and traumatized step-siblings. Sonya and Raskolnikov have known each other only a couple of days—though they learned of one another several days before the crime, they only met in person afterwards—but Raskolnikov has been kind to Sonya and her family, and she is already bound to him by deep gratitude, respect, and love. To her, and only to her, he is compelled to explain everything.

Raskolnikov's confession to Sonya is highly disjointed. He stops and starts over and over, with the expression "that's not it" characteristically signaling his failure each time to precisely and thoroughly express the reasons for his crime. This inadequacy of Raskolnikov's explanations does not arise from a lack of self-awareness—indeed, he is strikingly cognizant of the full complexity of his motives. Rather, the inadequacy arises from Raskolnikov's failure to acknowledge a basic feature of his psyche. His choices are determined through an internal dialogue among distinct, and at times opposing, "voices."[6] The internal debate among these voices cannot be resolved through rational deliberation because each voice expresses a unique sense of what is morally relevant. And the conscious choice to privilege one voice over others fails to silence the other voices or to create harmony within his conscience, so divided.

Thus, Raskolnikov's various attempts to characterize his motives in terms of a single privileged agenda fail because of an unacknowledged but inescapable complexity of his moral

[6] These "voices" do not have distinct personalities. Dostoevsky does not depict Raskonikov as suffering from dissociative identity disorder.

psyche, but his acknowledgment of the inadequacy of each of his explanations confesses precisely this complexity—altogether against his will. In this way, Raskolnikov's confession to Sonya illustrates all the major features of an original model of moral psychology that is implicit throughout Dostoevsky's major works—one that undermines ideas of moral psychology prevailing in Western theology and philosophy before Dostoevsky and that remains a highly relevant and untapped resource for the interdisciplinary science of moral psychology today. This dissertation is an attempt to explicate this Dostoevskian model of moral psychology and show why it is relevant to the current field of moral psychology.

Moral Polyphony

Philosophers typically organize their moral metaphysical philosophies around the idea of a single highest good or a hierarchical arrangement among goods, assuming that there really is a best way to behave in any situation. For Plato's Socrates, and for many philosophers and theologians after him, a failure to do what's best is simply a failure of rationality. If we were able to appreciate the best way to behave, we would have no incentive to behave in any other way.[7] While Aristotle, for his part, adopts a looser, "pluralistic" idea of goodness, his assumption that there is an ultimate goal of human life (*eudaimonia*: typically translated as "happiness"), as well as a relatively straightforward path to that goal, leads him to define moral virtues in such a way as to smooth over evident contradictions among different goods.[8] St. Thomas Aquinas expands Aristotle's teleology in accordance with his belief in a divinely ordered cosmos. Human virtues

---

[7] Plato. *'Protagoras' and 'Meno'*, trans. Robert Bartlett (Ithaca: Cornell University Press, 2004), p. 54-61 (*Protagoras* sections 352-358).

[8] Aristotle, *The Nicomachean Ethics,* ed. E. Capps, T.E. Page, & W.H.D. Rouse and trans. H. Rackham (London: William Heineman, 1934).

in this cosmos are fit into a harmonious metaphysical schema that excludes inconsistency *a priori*, by virtue of the fact that this cosmos is a divine creation.[9] Immanuel Kant's metaphysical moral system requires autonomous, rational beings to presuppose the existence of a single highest good, in accordance with which we can harmonize our will.[10] Only as such can his "categorical imperative" serve as a general principle for the determination of our (single, unambiguous) moral duty in each particular case. The founder of utilitarianism, Jeremy Bentham follows Socrates in calling the single highest good "pleasure,"[11] and while his philosophical heir John Stuart Mill modifies Bentham's view and adopts a pluralistic interpretation that embraces "higher" and "lower" forms of pleasure,[12] Mill's moral pluralism ultimately fits higher and lower goods into a fixed hierarchy. Any absolute contradiction between goods is here, as in all these examples, merely apparent.

Dostoevsky's work challenges all such ways of thinking about morality with his insight into the "polyphonic" structure of moral cognition. Polyphony is a style of music composition featuring multiple independent melodies, and it has also been deployed by the literary critic Mikhail Bakhtin as a metaphor for Dostoevsky's unique way of featuring a "plurality of

---

[9] Thomas Aquinas, *Summa Theologica*, trans. Fathers of the English Dominican Province (Ohio: Benziger Bros. Edition, 1947).

[10] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, ed. & trans. Allen Wood (New Haven: Yale University Press, 2002 [originally published in 1785].

[11] Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation* (1781; repr., White Dog Publishing, 2010), Chapter I: Of the Principle of Utility.

[12] John Stuart Mill, *Utilitarianism* (Kitchener: Batoche Books, 2014 [originally published in 1863]), Chapter II: What Utilitarianism Is.

independent and unmerged voices and consciousnesses" in his novels.[13] Whereas many other authors impose a "dialectical" structure onto their work, placing incomplete truths into the mouths of various characters only to synthesize them into a single overarching truth (often the author's own view), Dostoevsky endows his characters with voices that speak independently, sometimes even plausibly refuting his own convictions. By giving his characters competing moral ideas and impulses and generating manifestly unadjudicable debates within and among these characters, Dostoevsky initiates readers into an unfinishable conversation about (often moral) ideas.

According to Bakhtin, this unfinishable dialogue is the defining feature of Dostoevsky's polyphonic art form, but more important for our purposes is Bakhtin's suggestion that this feature derives from Dostoevsky's insight into human psychology.[14] Dostoevsky's polyphonic novel expresses the reality that cognition, especially moral cognition, is polyphonic. This dissertation is all about exposing the polyphonic reality of moral cognition and explaining it—showing where moral polyphony comes from and how it works in real time. Even though I will not be offering any metaphysical theory of morality, my Dostoevskian model of the polyphonic dynamics of moral cognition can be used to critique ancient and everyday assumptions about morality that many of us hold, consciously or unconsciously. One of my claims is that the reason we tend to think morality involves a simple dichotomous choice between right and wrong is that we aren't aware of the true complexity of our moral worlds.

---

[13] Bakhtin, *Problems of Dostoevsky's Poetics*, 6.

[14] Ibid., 40: "Dostoevsky could hear dialogic relationships everywhere, in all manifestations of conscious and intelligent human life; where consciousness began, there dialogue began for him as well."

Provisionally Defining Morality

Since the birth of moral psychology in the work of Jean Piaget, researchers have defined moral cognition circularly, treating various forms and outcomes of morality as if they were morality per se, and thus obscuring the idea of morality that all such moral forms and outcomes presuppose. For instance, we have tried to formally define moral cognition as cognition about moral values and principles. And yet, we have never had an adequate answer to the question: what makes a value or principle *moral,* as opposed to non-moral values and principles? Other definitions have emphasized the outcomes of cooperation, adaptive fitness, and developmental maturity. And yet, we have no good answer to the question of how to distinguish morality's contribution to cooperation, adaptive fitness, and developmental maturity from non-moral factors that also facilitate cooperation, adaptive fitness, and developmental maturity. This circularity and superficiality of our definitions causes us to ignore or misinterpret much that is decisive for moral cognition in everyday life.

A major contribution of this dissertation will be to address this deficiency and expose the real transcendental conditions of morality and moral cognition. I will do this work in chapter 3. For now, let us start with the most widely accepted operationalization of the moral domain: Elliott Turiel's distinction between moral norms and mere "conventions,"[15] where moral normativity is defined as a more serious and context-independent subdomain of normativity,[16] as

---

[15] Elliot Turiel, *The Development of Social Knowledge: Morality and Convention* (Cambridge: Cambridge University Press, 1983); Huebner, Bryce, James Lee, & Marc Hauser, "The Moral-conventional Distinction in Mature Moral Competence," *Journal of Cognition and Culture* 10.1-2 (2010): 1-26.

[16] The normative domain includes all standards for how things *ought* to be or behave. These can be aesthetic standards of beauty / ugliness, epistemic standards of truth / falsehood, instrumental standards of utility / uselessness, moral standards of good / bad or right / wrong, and more (appetizing / disgusting, interesting / boring, cool / uncool, and so on).

opposed to conventions, which are more context-dependent and may be violated with relative impunity. Turiel's operationalization does not give us definitive criteria for moral cognition—and we will see that the absence of such criteria has led the field into a protracted and theoretically confused debate—but it does give us a good hint, a place to start. Another hint is given by recent studies that indicate that one's moral character and widely shared moral beliefs are core to one's identity,[17] which I take as evidence that the relation between morality and identity is somehow fundamental. However, we should appreciate that in accepting these operationalizations, we have not really understood what morality and moral cognition are. Thus, in chapter 3, I will critique and round out these operational definitions with a phenomenological excavation of the transcendental conditions of morality and moral cognition.

Concluding Statement on the Polyphonic Model

My polyphonic model contradicts any psychological model or theory that fails to accommodate the real complexity of everyday moral cognition, along with all metaphysical systems that rely on such, shall we say, *homophonic* psychological assumptions. As such, moral polyphony is opposed not only to the moral "monism" of Socrates or Bentham but also to the "pluralism" of Aristotle or Mill. That is, my Dostoevskian model contradicts not only the idea that a single organizing principle, such as justice or happiness, can serve to satisfactorily orient moral education, but also the idea that there are several such principles, harmonious with each

---

[17] Geoffrey Goodwin et al., "Moral Character Predominates in Person Perception and Evaluation," *Journal of Personality and Social Psychology*, 106.1 (2014): 148-168; Nina Strominger & Shaun Nichols, "The Essential Moral Self," *Cognition*, *131* (2014): 159–171; Nina Strohminger & Shaun Nichols, "Neurodegeneration and Identity," *Psychological Science*, 26.9 (2015): 1469-1479.

other, that can perform this function together.[18] Moral polyphony is not simply a kind of pluralism of moral values or principles, opposed to the reductionism of metaphysical monism. My model arises from something deeper than all values and principles: our ontological relations to beings in the world. My central claim, which we see illustrated in Dostoevsky's novels, is that moral thinking involves an internal push and pull among shifting attachments and commitments, as we negotiate our sense of the value of other beings and our understanding of who and what we are in relation to the beings we care about. That is, the *ontological* complexity of our worlds gives rise to an irreducible complexity of moral cognition.

This polyphonic structure of moral thought is perhaps nowhere more vividly condensed than in the fifteen or so pages of *Crime and Punishment* wherein Raskolnikov confesses his crime to Sonya. Sonya's refusal to accept his initial explanations for why he did it induces Raskolnikov to give voice to the multifarious ontological attachments that influenced his action. The resulting image of human psychology contradicts assumptions about human nature that have been central to philosophical ethics in the West for thousands of years, flawed assumptions that continue to undermine theories in the field of moral psychology. The following reading of Raskolnikov's confession illustrates the key dynamics of this polyphonic moral-psychological model. It will serve as a reference for all subsequent analyses.

This scene takes place well after the murder, as Raskolnikov is trying to decide whether to kill himself or turn himself in to the authorities, or whether he might still be able to recover his strength and carry out his original plan. Leading up to his confession to Sonya, Raskolnikov is overcome by a sense of fatalistic powerlessness, similar to how he had felt shortly before

[18] This is not meant to imply that moral education is a hopeless endeavor. If there is an educational point here, it is that we should perhaps be more open to moral compromise.

committing the murder.[19] He doesn't understand why he feels so compelled, and we may ask ourselves the same question. So, I invite you to ponder why Raskolnikov feels the need to confess at all—to the authorities, to family and friends, and above all to Sonya in this moment—and what exactly he accomplishes through this confession. Allow this question to percolate as you read. We will return to it in the third chapter.

Raskolnikov's Confession

Raskonikov arrives at Sonya's directly after having saved her from an attempt by Luzhin, his sister's unsavory suitor, to frame Sonya for theft—threatening imprisonment and, at the very least, the thorough destruction of her reputation. If Sonya were to be incarcerated, her family would be without any source of income. Her stepmother would likely lose her mind (a process that has already begun), the children might starve, and Sonya's ten-year-old sister could herself be forced to resort to prostitution—a possibility that Raskolnikov has already taken pains to point out to Sonya.[20]

Raskolnikov fully appreciates the fragility of Sonya's position and her desperate desire to protect her family, and he opens the conversation with a thought experiment that, considering this, is barely hypothetical. What if Sonya had to choose either to allow her own family to be destroyed by Luzhin or to act violently against him? As readers are aware, one of Raskolnikov's motives for his crime had been the desire to save his own sister from a life of unhappiness with this very Luzhin, and we understand that Raskolnikov aims to strengthen the analogy between

---

[19] Dostoevsky, *Crime*, 62, 406.

[20] Dostoevsky, *Crime*, 321, 407.

Sonya's situation and his own, arousing empathy for his motives and undermining any moral condemnation of his crime in advance of his confession.

Raskolnikov has laid the groundwork for this analogy in a prefatory comment, reminding Sonya of her choice to become a prostitute to save her family.[21] According to Sonya's own value system, she is not morally permitted either to let her family starve or to live a life of prostitution, and yet the harsh realities of life have forced her to make this choice anyway. Why should Sonya be willing to profoundly violate herself but stop short at violence against an evil man? Thus, while holding back his confession for the moment, Raskolnikov begins to draw an analogy between his crime and Sonya's tragic choice, implicitly framing his own act as an analogous moral dilemma: "if all this was suddenly given to you to decide […] How would you decide which of them [Luzhin or Sonya's family] was to die?"[22] But Sonya rejects such a choice, rendering Raskolnikov's careful analogy irrelevant before it can be made explicit: "I cannot know divine Providence…And why do you ask what cannot be asked?"[23] Without seeking to, Sonya has dashed Raskolnikov's hope of justifying his action to her, and Raskolnikov abandons this first confession attempt with an apology that is totally enigmatic, given his failure to confess: "I told you yesterday that I would not come to ask forgiveness, and now I've begun by almost asking forgiveness…I was speaking about Luzhin and Providence for my own sake…I was seeking forgiveness, Sonya..."[24] If "forgiveness" means Sonya's moral approval of his deed, Raskolnikov can forget it.

---

[21] Dostoevsky, *Crime*, 406.

[22] Ibid., 408.

[23] Ibid.

[24] Ibid.

Raskolnikov struggles to speak. Sonya's refusal to consider condoning murder under any circumstances stands as a preemptive condemnation of his crime, putting him on the defensive: "And suddenly a strange, unexpected feeling of corrosive hatred for Sonya came over his heart."[25] But Raskolnikov's defensive hatred is dissipated by his contact with "her anxious and painfully caring eyes,"[26] a personal connection that undermines his impulse to aggressively justify himself before her. He can see that Sonya's moral judgment of his action is not rigidly determinative of her assessment of his person, that however wrong she considers his crime to be, she cannot see him as a Luzhin—unscrupulous and morally despicable—and this realization softens his heart. Raskolnikov had been using Sonya as a proxy—his attacks against her religious faith are also attacks upon some of his own moral feelings, which he sees as weakness. But in the moment of their encounter he recognizes the opposition between his need for self-justification and Sonya's need to maintain faith in a moral ideology that defines and sustains her existence. If he views the idea of divine providence as pathetic and irrational, he can appreciate nonetheless that it is Sonya's greatest source of hope. In this face-to-face encounter, Raskolnikov's focus suddenly shifts away from Sonya's moral ideology and onto her person, and he momentarily abandons the project of rational justification, as aggression gives way to empathy: "Only why did I come to torment you?"[27]

Abandoning his plan of rationally justifying his action to Sonya, Raskolnikov also forfeits the hope of using his confession as a vehicle for self-justification. Still, he is by no means

---

[25] Ibid.

[26] Ibid., 409.

[27] Ibid.

prepared to lower himself in humility and admit to wrongdoing. But as this door closes, Sonya's compassion and esteem for Raskolnikov opens a new way for him to confess the crime while maintaining his dignity. If her respect for him is not entirely dependent upon her opinion about the morality of his action, he doesn't have to choose either to mount a proud self-defense or to offer a humble renunciation. He can just stand in his truth. Thus, when the confession finally comes, it does not take the form of a verbose explanation characterized by either self-righteousness or regret. It comes rather as a wordless and utterly vulnerable gaze.

The two have reached such a state of empathic entanglement that affective states pass contagiously between them. Sonya feels Raskolnikov's suffering as if it were her own, and Raskolnikov is equally affected by Sonya's feelings. As the wordless confession sinks in, Sonya backs away from Raskolnikov in terror, and this terror "communicated itself to him: exactly the same fright showed on his face as well."[28] This is a hard truth to express and to receive—Lizaveta had been one of Sonya's only friends. Yet, in spite of her fear and devastation, Sonya's response is compassionate: "what have you done to yourself!"[29]

Raskolnikov has been living in isolation, having set himself above and against everyone around him, and he desperately needs this personal connection: "A feeling long unfamiliar to him flooded his soul and softened it all at once. He did not resist: two tears rolled from his eyes and hung on his lashes."[30] And yet, while confessing to a crime against the law (in both a legal and a moral sense), Raskolnikov is not ready to confess that his action is ultimately *wrong*. For Sonya, the solution is devastatingly difficult, but unquestionable. She does not suspect how far

---

[28] Ibid., 411

[29] Ibid.

[30] Ibid., 412.

Raskolnikov is from accepting her moral premises, and, because his fierce but fragile pride is so foreign to her, she has no sense of the volatile state of his ego in that moment: "Again she embraced him. 'I'll go to hard labor with you!' He suddenly seemed to flinch; the former hateful and almost arrogant smile forced itself to his lips. 'But maybe I don't want to go to hard labor, Sonya'."[31]

Faced with this intractability, Sonya seeks to understand Raskolnikov's motives. His defenses are back up, but his strategy has changed. Instead of seeking to justify himself before her, he now tries to prevent further probing. He feels ashamed of his early attempt to manipulate Sonya into condoning the crime, and now he does an about-face, construing his action so as to maximize his blameworthiness and destroy any possible analogy to her situation. Thus, he responds to the question "why?" with put-on psychopathic levity: "To rob her, of course."[32]

Taking Raskolnikov at his word for the moment, Sonya tries to ameliorate his guilt by suggesting that he robbed out of hunger and to help his mother. In bringing up such extenuating circumstances, Sonya touches on another key category of Raskolnikov's motives, though in doing so she does not guess the rigor of his meditations upon these matters. He has privately enumerated many versions of what philosophers call "consequentialist" or "utilitarian" ethical positions—different variations on the basic idea that the "ends justify the means." He has obsessed over the compromised dignity and well-being of his mother, his sister, and himself—the latter less in terms of his basic need for food and shelter (though he was in real danger of homelessness and starvation) and more in terms of his opportunities to do important work, change the world, and make a name for himself. He has also been very deliberate in choosing his

[31] Ibid.

[32] Ibid.

victim, enumerating to himself compelling reasons why the world would be a better place without the extortion practiced by Alyona Ivanovna, and why his use of her money would benefit people in the world infinitely more than her miserly hoarding of the money in her lifetime and her absurd plan for its use after her death.[33] In view of all such considerations, part of Raskolnikov is convinced that the murder of the pawnbroker is morally justifiable—indeed, a consequentialist could hardly ask for a more airtight case in favor of the deed. And Sonya is correct in positing some consequentialist considerations as motivating factors for his crime, though she does not see these considerations as rational justifications. But there is some completely different reason for Raskolnikov's unhappiness, stemming from the fact that all these extenuating considerations are not quite *it*. It would be easier, Raskolnikov feels, if it were something that simple: "if I'd killed them only because I was hungry […] I would now be…happy!"[34] Sonya's attempt to offer ameliorative context for Raskolnikov's crime had not been intended as a moral justification for the murder, and the very idea is offensive to her sensibility and inconsistent with her sense of Raskolnikov's character. Nevertheless, she understands the crucial part of his strange claim—that his consciousness of some altogether different motive is responsible for the depth of his present torment.

Just explaining this other motive to Sonya involves a kind of metaphysical and psychological aggression, since Raskolnikov must express his philosophical dismissal of the Christian ideology that gives Sonya's existence meaning and the self-abnegation she embodies,

---

[33] As an opportunistic moneylender, Alyona has preyed upon the poor in her lifetime, amassing a fortune, which she hoards with the sole purpose of paying clergy to intervene for her soul after her death, planning (from Raskolnikov's point of view) to buy her way into heaven with money stolen from the poor. See Dostoevsky, *Crime*, 64.

[34] Dostoevsky, *Crime*, 413-414.

and he must therefore express a certain contempt he harbors for her moral ideals and her basic way of being in the world. Shying away from such aggression, Raskolnikov begins not with an explicit elaboration but with an enigmatic emblem: "I wanted to become a Napoleon."[35] With this emblem, he has said *all*, but if Sonya wants to understand what it means, she will have to ask him to explain. In this way, Raskolnikov distances himself from the effects of his idea upon Sonya, bypassing his own feelings of empathy and responsibility for her and giving them both over to the fateful momentum of the conversation.

Seeing, however, at her first expression of confusion that he cannot explain his idea to Sonya in this cryptic manner, Raskolnikov reverts mechanically back to his original design, seeking to flesh out the consequentialist justification for the crime that he had earlier failed to make explicit. This time he pits a straightforward utilitarian logic against the sacred law against murder, explaining his action in a way that is maximally sympathetic. He needed the money to save his mother and sister from a tragic fate and set up a career for himself, which would allow him to do a great deal of good in the world.[36] This style of justification is distasteful to Raskolnikov, however, and he recites it without conviction. He doesn't really want to evoke pity. He doesn't want to have to calculate utility. He wants someone to endorse his action without reservation, and this is never how such utilitarian decisions are affirmed. Thus, at Sonya's first sign of disbelief, Raskolnikov immediately capitulates: "You can see for yourself that's not it!…yet it's the truth, I told it sincerely!"[37]

---

[35] Ibid., 415.

[36] Ibid., 415-416.

[37] Ibid., 416.

Having abandoned this sincerely spoken truth as altogether "not it," Raskolnikov shifts the blame back onto himself. From the beginning, however, his tone is internally inconsistent, as he distances himself from the seeming self-deprecation of the gesture: "Better…suppose (yes! it's really better this way), suppose that I'm vain, jealous, spiteful, loathsome, vengeful, well…and perhaps also inclined to madness."[38] Despite this "suppose," there is some truth to this confession, and Raskolnikov goes on to offer crucial information about his psychological condition in the months leading up to the crime: "I could have earned enough for boots, clothes, and bread myself; that's certain! There were lessons; I was being offered fifty kopecks. […] But I turned spiteful and didn't want to."[39]

Against what did Raskolnikov turn spiteful? From the beginning of the novel, readers have seen him struggling with the stupidity, cruelty, and injustice all around him in 19th century Petersburg. It seems almost a general law that the innocent and vulnerable suffer, the brilliant and original fail, and the mean and stupid get their way, and Raskolnikov's spite springs at least partly from moral and aesthetic outrage at this perceived reality. Hence his inconsistent tone: even as Raskolnikov presents his spitefulness as a weakness of character that stops him from taking positive steps to save himself and his family, another voice in his psyche interprets things differently. His is a noble spite, born of outrage over injustice. This other voice interprets Raskolnikov's spite as a sign of his superior moral sensitivity and the uncompromising nature of his conscience.

---

[38] Ibid., 417

[39] Ibid.

Witnessing the daily exploitation of the most vulnerable, and seeing how such exploitation is so often justified through absurd casuistry, the only thing Raskolnikov feels he can affirm is an act of aggression against the source of it all: the system itself. To embrace Sonya's Christian morality and pursue a life of humble service would perhaps only enable the perpetuation of exploitation, and Raskolnikov feels that the compassionate self-sacrifice of Sonya and others like her is implicated in the injustice of this system.[40] So he rejects the sacred values of his society, which seem so susceptible to distortion and so disconnected from actual human flourishing, and he seeks to liberate himself through destruction. The miserly pawnbroker Alyona Ivanovna seems to exemplify the absurdity of 19th century Petersburg society, with its mixture of capitalistic viciousness and religious casuistry. Through his violence against what he sees as an emblem of human stupidity and corruption, Raskolnikov can express his aggression towards society at large.

The act of lashing out against absurdity is, for Raskolnikov, ripe with meaningfulness. One who would dare, for the sake of an idea, to flout the laws that prop up an unjust social order, could create meaning through destruction, setting himself above and against these grotesque dynamics and imbuing his character with a certain sublime dignity. Thus, in conceiving such an aggressive response to a repulsive reality, Raskolnikov merges his moral and aesthetic sensibilities and seeks to respond to the situation in which he finds himself with an artistic balance of sensitivity, realism, and boldness:

> […] then a thought took shape in me, for the first time in my life, one that nobody had ever thought before me! Nobody! It suddenly came to me as bright as the sun: how is it that no man before now has dared or dares yet, while passing by all this absurdity, quite simply to take the whole thing by the tail and whisk it off to the

[40] Ibid., 43-44.

> devil! I…I wanted to dare, and I killed…I just wanted to dare, Sonya, that's the whole reason![41]

Here he expresses his idea in terms of a daring action, but we have already seen Raskolnikov express the idea in terms of a person: a "Napoleon." The action signifies a personal quality, and Raskolnikov's desire to commit a daring act arises from his desire to be a daring person. Readers understand what this means better than Sonya because we are privy to an earlier conversation between Raskolnikov and the cunning detective Porfiry Petrovich, wherein the latter goads Raskolnikov into expounding upon his ideal. A "Napoleon" represents for Raskolnikov a superior type of person: someone who can rise above the moral laws that the mass of humanity must obey, and instead act out of an affirmation of their own superiority. A "Napoleon" is thus, for Raskolnikov, someone who has a kind of *moral right*[42] to transgress against the moral law.

In this early conversation, Raskolnikov and the detective discuss an essay that Raskolnikov had submitted for publication while still in school but that has only recently been published, unbeknownst to him, in a different journal. In the article, Raskolnikov illustrates his ideal by dividing the human race into two types: the ordinary and the extraordinary. Ordinary types are the mass of humankind—people who are in touch with the moral demands of the present and who believe in the sacred values of their socio-historical moment. According to Raskolnikov, it is fitting for such people to adhere to their moral law and to be punished when they transgress. But there is another type, which is exceedingly rare. Such "extraordinary"

---

[41] Ibid., 418.

[42] We will see in chapter 3 precisely how this is a *moral* right.

people, who "have the gift or talent of speaking a *new word,*"[43] have a right, even a duty, to transgress against the morality of their moment for the sake of a creative vision of the future with which they are uniquely endowed. They have a right to transgress because their actions are infinitely more important and perhaps even more beneficial to the world than the actions of normal people. They see farther into the future and weigh good and bad in a greater context, whereas the mass of humankind is not capable of such vision and must be made to submit blindly to the moral law of their day. Moreover, even if the mass of humanity would not agree that their idea of the future is really superior, such extraordinary "Napoleons" would pay this no mind, since they know that their understanding is different and better than that of the masses. Thus, there is a kind of consequentialist (though not utilitarian)[44] justification for the transgressions of these rare humans. Their "ends" are indeed used to justify their "means," but the criteria by which ends and means are evaluated remain subjective. When it comes to distinguishing right from wrong, what ultimately matters to such "Napoleons" is their unique vision of the future.

Sonya has not heard this argument and would not accept it if she had, and she opines that Raskolnikov's aggression against "all this absurdity" derives from an evil source—the devil.[45] Surprisingly, Raskolnikov agrees with this, although his concept of "devil" is altogether different from hers. For Sonya, the devil is a living being, an evil spirit who lies and tempts. There are several possibilities for what "devil" could mean to Raskolnikov, who is an atheist: the self-aggrandizing tendency of his own mind, mysterious forces of fate, situational features of the

---

[43] Dostoevsky, *Crime*, 260.

[44] Utilitarians in the tradition of Jeremy Bentham only consider consequences morally relevant insofar as they relate to the pleasure and pain of sentient beings.

[45] Dostoevsky, *Crime*, 418.

environment. Whatever Raskolnikov means by "devil," it is something outside of his will, some external force that drags him along helplessly in its wake. For Raskolnikov, this idea is neither comforting nor mitigating; it strips him of the dignity he desires above all else. Porfiry Petrovich has already provoked Raskolnikov by suggesting that he had been misled by his own ego into imagining himself to be "a Napoleon,"[46] and this thought torments Raskolnikov, a wound to his pride that will not heal on its own. He feels driven to confess this shameful truth, and so he grants Sonya's point, but his response is faithful to his own concept of "devil": "I know myself that a devil was dragging me […] do you really think I didn't at least know, for example, that since I'd begun questioning and querying myself: do I have the right to have power?—it meant that I do not have the right to have power?"[47]

We must take a few moments to unpack the remarkable ethical theory implicit in this statement. With this admission, Raskolnikov has already begun to express a more radical feature of his idea, which he had withheld in his early encounter with the detective. Whereas in that conversation Raskolnikov had emphasized the suffering of "truly great men"[48] and suggested that Napoleonic types must carry their conscience like a great and unshakable weight, now he expresses a more hopeful vision. Perhaps such a transcender of the sacred moral order could shake off their moral shackles and become weightless. Indeed, in terms of conscience, perhaps "he who dares the most will be the lightest of all!"[49] Perhaps in being so daring, he wouldn't

---

[46] Ibid., 265.

[47] Ibid., 418-419.

[48] Ibid., 264.

[49] Ibid., 418.

need to justify himself in any way, but might effectively stick out his tongue at the empty casuistry of *all* preordained morality:

> I longed to shake it all off my back: I wanted to kill without casuistry, Sonya, to kill for myself, for myself alone! I didn't want to lie about it even to myself! It was not to help my mother that I killed—nonsense! I did not kill so that, having obtained means and power, I could become a benefactor of mankind. Nonsense! I simply killed—killed for myself, for myself alone—and whether I would later become anyone's benefactor, or would spend my life like a spider, catching everyone in my web and sucking the life-sap out of everyone, should at that moment have made no difference to me!"[50]

Above all, the murder wasn't for money, he says, but to find out "whether I was a louse like all the rest, or a man? Would I be able to step over, or not! Would I dare to reach down and take, or not? Am I a trembling creature, or do I have the right…"[51] The murder was thus a kind of test, a way for him to find out if he was ordinary or extraordinary.

With this confession, Raskolnikov abandons any consequentialist justification of his crime and for the first time expresses the most radical dimension of his idea. When earlier he confessed to Sonya that he "wanted to become a Napoleon," it was an acknowledgment both of the more superficial presentation he had given to Porfiry Petrovich and of this deeper layer of his idea. In either case, a "Napoleon" is a rare type of person who has, on account of his superior qualities, a right to transgress against the moral law of his socio-historical moment. However, whereas the first presentation relies upon a kind of consequentialist justification for the crime,

---

[50] Ibid., 419.

[51] Ibid.

the deeper layer of Raskolnikov's idea involves an original, proto-existentialist form of justification—something that "nobody had ever thought before."[52]

This existentialist justification proceeds from a presupposition that god and morality are ideas that only *exist* as they are manifest in the experience of those who believe in them. Freeing oneself from these ideas means freeing one's conscience from subordination to them and thereby freeing oneself from guilt over one's transgressions. The feelings of the transgressor thus function as a direct way for them to know if they are bound by the moral law that binds normal people or if, on the other hand, they have transcended this law: if, in transgressing, they feel not guilt but instead some unquestioning conviction of the rightness of their action, then they have performatively transcended the moral law that enslaves everyone else. Within their reality, their experience of being in the world, they effectively *have* a right to transgress. And this right has been confirmed through their transgressive act. Thus, in committing the act, such a person may conceivably not care "whether I would later become anyone's benefactor, or would spend my life like a spider." The point is not the outcome but the experience itself, since, in the realm of ideas, experience *is* reality.

This possibility of transcendence is what Raskolnikov means by lightness, his secret hope. It must be sharply distinguished from a mere absence of conscience, the psychopathic levity that he feigned earlier in his confession. Anachronistically, we might call Raskolnikov's ideal Nietzschean. The values that such "Napoleons" create are themselves sacred. Such "extraordinary" types have an unshakeable faith in the meaningfulness of their activity, and they may even consider it their duty to transgress against the common moral law, whether for the sake of their creative vision of the future or for the experience of moral transcendence itself. Thus,

---

[52] Whether or not this idea is original, Raskolnikov's assertion of the originality of his idea is self-serving, as it places him in the class of "Napoleons," who have some "new word" to speak.

their transgressions do not express an absence of moral sensitivity but a moral achievement—a re-conscription of the conscience in service of a new goal—what Nietzsche would later call a "transvaluation of values."[53]

Sonya will never concede any moral justification for murder, and by the end of their conversation, Raskolnikov no longer holds out hope for this concession. He only wants to do penance by confessing the most humiliating thing of all: that in conceiving his bright vision he had been deceived by the self-aggrandizing tendency of his mind—a double-edged "devil": "I wanted to prove only one thing to you: that the devil did drag me there then [to commit the murder], but afterwards he explained to me that I had no right to go there, because I'm exactly the same louse as all the rest!"[54] This "devil" now reveals to Raskolnikov that he too is one of those detestable moral bugs—an ordinary person. When it comes down to it, he is not daring enough to leap free of the moral confines of his historical moment and carry forth a bold artistic vision. Instead, he must adopt the same values that all the ordinary people have and cannot help having. Indeed, he now confesses, part of him knew it all along. He knew he didn't have it in himself to kill, knew that the deed would be a disaster for him—so why did he do it? Ultimately, he now says, the deed was an act of aggression against himself, against that part of him that was always ordinary and always knew it: "Was it the old crone I killed? I killed myself, not the old crone! Whopped myself right then and there, forever!…And it was the devil killed the old crone, not me…"[55]

---

[53] E.g. Friedrich Nietzsche, *Beyond Good and Evil,* trans. Helen Zimmern (New York: Barnes and Noble, 2007), aphorism 203, p. 91.

[54] Dostoevsky, *Crime*, 419.

[55] Dostoevsky, *Crime*, 420.

Temporarily broken by his confession, Raskolnikov meekly asks Sonya what he should do. She does not need a full account of his philosophical vicissitudes to know how to answer him—he must repent and renew his bonds with the earth and the society of which he forms a part: "Go now, this minute, stand in the crossroads, bow down, and first kiss the earth you've defiled, then bow to the whole world, on all four sides, and say aloud to everyone: 'I have killed!' Then God will send you life again."[56] This solution to moral nihilism, to which Dostoevsky returns in all his major works, stems from his insight into the polyphonic structure of the human mind. One must give up the attempt to achieve moral coherence through philosophical ratiocination—an attempt that is doomed in advance by the irreducible complexity of moral cognition—and return through repentance to a state of philosophical innocence, reconnection with society and nature, and mystical faith.

Moral Polyphony Revisited

I take Raskolnikov's emotional conclusion to be entirely sincere. However, even in the final and most honest iteration of his confession to Sonya we do not find the whole truth. The truth that he is confessing is complex and polyphonic. The murder wasn't simply an act of aggression against his own weakness. Raskolnikov had also been moved by practical consequentialist considerations: he *had* thought about how much more good he could do with the pawnbroker's money than she ever would; he *did* really want to help his mother and sister and himself. Raskolnikov was *also* moved by a kind of aesthetic-moral outrage at the absurdity around him, especially as exemplified by the viciousness and stupidity of Alyona Ivanovna and

---

[56] Ibid.

Pyotr Petrovich Luzhin. He *did* imagine, in view of this, that his crime might be noble and, in a quasi-moral sense, better than the meek self-sacrifice of someone like Sonya. And he *also* hoped that in committing his crime he might transcend the narrow morality of his socio-historical moment and realize his own extraordinary mission, manifesting the broader and nobler, but also infinitely lighter conscience of a "Napoleon."

None of these explanations is sufficient by itself, and Raskolnikov's admission of the inadequacy of each attempt expresses the complexity of his polyphonic conscience. But, this complexity being granted, it would be reasonable to question whether we can learn much about the psychology of ordinary people from this fictional depiction of a pathologically conflicted person, who also happens to be a murderer. Raskolnikov's name stems from the Russian word for "schism,"[57] and he is indeed schismatic in the sense that his feelings, thoughts, actions, and even his wildly fluctuating bodily states constantly express conflict between opposing impulses and convictions. And this dynamic gives rise to a number of contradictory personal qualities. Raskolnikov possesses a brilliant rational mind, and yet his thinking falls constantly into paradox and confusion. He has an overwhelming desire to do something important, and yet he is incapable of resolute action. He constantly vacillates between love and hatred for those closest to him and cannot decide whether to feel compassion or disgust for everyone else. And yet, in spite of the unique dynamism of his internal contradictions, I maintain that the basic structure of Raskolnikov's thought is a fitting representation of moral thinking in general.

The typicality of this psychic structure begins to become evident when we look at the elements of Raskolnikov's internal dialogue one at a time: ambition, a need for self-esteem, empathy, a sense of justice, disgust, a practical-rational faculty, a legalistic-moral faculty, guilt, a

[57]Dostoevsky, *Crime*, from Richard Pevear's introduction.

need for intimacy, a sense of the sacred, an intellectual conscience, shame, aggression, a need for forgiveness, a sense of responsibility for his own future and those closest to him, all his various social and extra-social attachments, and so on. None of the elements on its own is uniquely Raskolnikovian, and most are commonly present to some degree in most people. Moreover, the internal vicissitudes we see dramatized in Raskolnikov's person follow plausibly from the juxtaposition of these ordinary elements in his unique circumstances. His rational justification of the crime, and all the good he imagines might come of it, cannot stop him from feeling horrible about the murder itself; and so, no matter what he does or does not do, he is bound to wind up in a cul-de-sac of regret. His proud intellect is a two-faced "devil," building his hopes with a sublime vision of human possibility only to cut him down to the size of a mere "louse." His desires for forgiveness, self-esteem, and justice generate contradictory demands that render him incapable of resolute action. His ambition to manifest an "exceptional" subjectivity runs up against his attachments to his loved ones and the broader society. His empathic connection to Sonya is in tension with his brittle egoism, his intellectual conscience, and his aggression towards her religious ideology, and this tension contributes to the constant emotional vacillations of his confession. These dynamics make enough sense that we shouldn't need to ask where Raskolnikov's schismatic nature comes from. Instead, we should wonder why most people, who possess the same basic psychological elements as Raskolnikov, are *not* so extremely and pathologically conflicted most of the time.

Looking at Dostoevsky's work more broadly, one encounters an insistent claim about the nature of Raskolnikov's pathology and at least a partial answer to the question why most people are not perpetually in a state of internal moral conflict. Several of Dostoevsky's characters suffer in a similar way, including the unnamed protagonist of *Notes from the Underground* (henceforth

the "Underground Man"), Ivan Karamazov, and Raskolnikov. The Underground Man gives this kind of suffering an explicit diagnosis when he claims to be "ever aware of the great number of completely conflicting elements within"[58] himself, and says that he suffers from too much "consciousness."[59] The question, then, is why all of us do not suffer constantly from "consciousness"—from guilt and indecision caused by conflict among the various elements that feed into moral thought. Is it because our psyches do not contain conflicting elements? Or might it simply be that we are not often conscious of the conflicts between these elements of our polyphonic psyches, because we are less driven than Raskolnikov to bring them into critical dialogue?

Jonathan Haidt's Social Intuitionist Model of moral psychology proposes that most people tend to make moral decisions rapidly and automatically most of the time, not worrying overmuch about the internal consistency of their beliefs and behaviors until they are challenged to defend them.[60] Thus, per Haidt, it is primarily critical dialogue—most often with other people but also within our private thoughts—that generates our attempts to justify our actions in a principled way. Absent this, we can get along without any awareness of the many moral contradictions we manifest daily in thought and action. Moral thought, according to this model, is strategic and motivated, not impartial and dispassionate. Reading Dostoevsky's work with an eye toward moral psychology, I find support for this thesis and an additional suggestion that in many cases such dialogues can never be resolved because they express a confrontation between

---

[58] Fyodor Dostoevsky, *Notes from Underground*, trans. Jane Kentish (New York: Oxford University Press, 1991), 8.

[59] Ibid., 10.

[60] Jonathan Haidt, "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108 (2001): 814-834.

essentially incompatible commitments. Because of the nature of moral thought, perfect moral harmony is not possible in this world. Because of this unresolvability of much moral deliberation, when we insist on harmonizing all of our moral beliefs in accordance with a single goal or standard and pursue this internal consistency with all rigor, we are effectively beating our heads against a wall.

Thus, one response to the objection that Raskolnikov's schismatic nature cannot tell us anything about functional human psychology is that the main reason typical people are not as pathologically conflicted as Raskolnikov is simply that we are not motivated to push our ethical theorizing into this pathological territory of beating our heads against a wall. We blindly assert our own righteousness. Or we ask for forgiveness and try to make amends. Either tactic can preserve our false sense of moral coherence, and both strategies work by pulling us up short of a deep dive into the true moral complexity of the matter. Raskolnikov is different because he dives so much further into this complexity. But he is like us in the way he dives, assuming, in spite of all his skepticism and iconoclasm, that there is a moral truth to be found at the bottom—that he either had "the right" to do what he did, or that he did not have the right. The truth is that there is no single moral truth but multiple voices, forcing choices that are often imperfect but that we must make nonetheless. Raskolnikov undergoes such a choice like everyone else. In spite of his hyper-awareness of everything that is at stake, he really is "exactly the same louse as all the rest."[61]

This Dostoevskian interpretation of Raskolnikov's pathology being understood, we are prepared to appreciate why Raskolnikov is not merely suitable but ideal for illustrating a Dostoevskian model of moral psychology. Raskolnikov dramatizes a dynamic that remains

[61] Dostoevsky, *Crime*, 419.

relatively inert in most people most of the time. His travails are thus like earthquakes and volcanoes hinting at the unseen fault lines of functional human psychology. Raskolnikov's conflicts are signs indicating where we should introspect if we want to perceive our own internal schisms. With characters like Raskolnikov, Dostoevsky thereby initiates us into a phenomenological mode of investigation of human nature, wherein the ordinary features of our minds are understood through reflection upon key limit-experiences. With the richness of his illustrations, Dostoevsky provides something that is lacking in both psychological and philosophical accounts of moral cognition.

Such a phenomenological approach is a valuable and under-utilized guide for empirical psychological research. Moral psychologists cannot escape phenomenological interpretation, and if we try to avoid it, we simply do it badly. This is the case in the first place because our domain of inquiry—"morality"—is phenomenologically defined. Typical operationalizations of the term do not escape this, because they simply average over the phenomenological interpretations of participants. If we know that morality is a special domain of normative judgments or that shared moral beliefs are repositories or indicators of moral-communal identity, we only know these things from asking people what they think, from mining in this way their experiences, and their interpretations of experience—in short, their folk-phenomenologies. Such operational definitions are useful, but we cannot stop with the phenomenology here, having staked out our domain and determined what is to be measured. Too often the subject of inquiry in the field of moral psychology is a shallow, uncomplicated interpretation of morality and moral cognition. By contrast, Dostoevsky's work offers a rich phenomenological vision of the domain of morality and the psychological dynamics involved in moral judgment. As such, his work stands as a

challenge to moral-psychological science. Can our models and theories accommodate *that*? If not, then what are we doing?

I don't expect you to be convinced about the usefulness of literature for the field of moral psychology just yet. In the following chapters, I bring my Dostoevskian model into dialogue with empirically grounded research and theory in the field of moral psychology, and I take Raskolnikov's confession as a challenging case for my polyphonic model to explain. The important test, as I see it, is whether my model leads to a more adequate interpretation than other models and theories—both of existing scientific research and of lived experience, such as that illustrated in Dostoevsky's work. I will leave this judgment to the reader.

Chapter II: Evolution of Morality

Chapter II Contents:

"There's no one here who is worth such words!" Aglaya burst out. "No one, no one here is worth your little finger, or your intelligence, or your heart! You're more honest than all of them, nobler than all of them, better than all of them, kinder than all of them, more intelligent than all of them! There are people here who aren't worthy of bending down to pick up the handkerchief you've just dropped…Why do you humiliate yourself and place yourself lower than everyone else? Why have you twisted everything in yourself, why is there no pride in you?"[62]

Introduction: Evolution of Socio-Moral Psychological Features

When imagining how our moral psychology may have evolved, theorists often take a comparative approach, looking at similarities and differences between humans and other social animals. As highly social primates, humans are specially adapted to engage in various forms of cooperation and socially mediated competition. We share many features with other social primates, including hierarchical tendencies and the ability to cooperate. But humans also stand out among primates in numerous ways. For us, hierarchical status is determined not just by physical dominance or birthright, but also by prestige, which may be gained by people with desirable personal qualities or resources.[63] Like other primates, our pursuit of hierarchical status is competitive; but it is also a pursuit of social connection, and our deference to prestigious individuals may help streamline cultural learning.[64] Humans develop very slowly and benefit greatly from cooperation, facilitated by our capacities for language and culture.[65] Supporting this slow developmental process—which gives us time to learn the accumulated knowledge of our

---

[62] Fyodor Dostoevsky, *The Idiot*, trans. Richard Pevear and Larissa Volokhonsky (New York: Vintage Books, 2003), 342.

[63] Joseph Henrich & Francisco Gil-White, "The Evolution of Prestige: Freely Conferred Deference as a Mechanism for Enhancing the Benefits of Cultural Transmission," *Evolution and Human Behavior* 22.3 (2001): 165–96.

[64] Ibid.

[65] Natalie Henrich & Joseph Henrich, *Why Humans Cooperate: A Cultural and Evolutionary Explanation* (Oxford: Oxford University Press, 2007).

social group—human fathers, grandparents, and other group members contribute much more to child provision and care than non-mothers of any other primate species[66]; and unlike the females of other primate groups, women live for many years after the cessation of reproduction,[67] supporting younger mothers and grandchildren and providing vital cultural and ecological knowledge, resources, and leadership.[68] Finally, management of this complex social life is facilitated by our sensitivity to social norms and our willingness to enforce these norms, even at personal cost.[69] This, in skeleton, is what human sociality looks like, from a comparative evolutionary point of view.

Evolutionary theorists of moral psychology tend to see moral cognition as a subspecies of social cognition. While I will show in chapter 3 that moral cognition exceeds the boundaries of social cognition, I acknowledge that there is a great deal of overlap between these domains. After all, humans make great use of special moral-cognitive abilities in navigating our social world. We are strikingly responsive to signals of personal qualities like moral character,[70] especially in

---

[66] Hillard Kaplan, et al., "The Evolutionary and Ecological Roots of Human Social Organization," *Philosophical Transactions of the Royal Society B* 364 (2009): 3289-3299.

[67] Susan Alberts, et al., "Reproductive Aging Patterns in Primates Reveal that Humans are Distinct," *Proceedings of the National Academy of Sciences of the United States of America* 110.33 (2013): 13440–13445.

[68] Darren Croft, et al., "The Evolution of Prolonged Life After Reproduction, *Trends in Ecology & Evolution* 30.7 (2015): 407 – 416; Kristin Hawkes, et al., "Grandmothering, Menopause, and the Evolution of Human Life-Histories," *Proceedings of the National Academy of Sciences U.S.A*. 95 (1998): 1336–1339.

[69] Herbert Gintis, "Strong Reciprocity and Human Sociality," *Journal of Theoretical Biology* 206 (2000): 169-179; Herbert Gintis et al., "Strong Reciprocity and the Roots of Human Morality," *Social Justice Research* 21 (2008): 241-253.

[70] Geoffrey Goodwin, et al., "Moral Character Predominates in Person Perception and Evaluation," *Journal of Personality and Social Psychology* 106.1 (2014): 148-168.

potential mates.[71] We are great at inferring what others are thinking and feeling, a skill that comes in handy when we wish to help[72] or deceive.[73] We also have features that help us detect liars and cheaters—including a sensitivity to normative rules[74] and an ability to use gossip to find out about each other's behavior.[75] We feel empathy for those to whom we are close—especially our children—and this helps motivate care and cooperation.[76] We have a deeply ingrained sense of fairness, which guides us as we form and maintain exchange-relations.[77] And we are intuitively sensitive to injustice, meanness, and disloyalty (among other things), having strong feelings that motivate our responses to such socially destructive behaviors.[78]

---

[71] Kristen Hawkes and Rebecca Bliege Bird, "Showing off, Handicap Signaling, and the Evolution of Men's Work," *Evolutionary Anthropology* 11 (2002): p. 64; Geoffrey Miller, "Sexual Selection for Moral Virtues," *The Quarterly Review of Biology* 82.2 (2007): 97-123.

[72] Carolyn Zahn-Waxler, et al., "Development of Concern for Others," *Developmental Psychology* 28.1 (1992): 126-136.

[73] Victoria Talwar and Kang Lee, "Social and Cognitive Correlates of Children's Lying Behavior," *Child Development* 79.4 (2008): 866-881; Xiao Ding, et al., "Theory of Mind Training Causes Honest Young Children to Lie," *Psychological Science* 26.11 (2015): 1812–1821.

[74] Fabrice Clement, et al., "Social Cognition is not Reducible to Theory of Mind," *British Journal of Developmental Psychology* 29 (2011): 910-928.

[75] Robin Dunbar, "Gossip in Evolutionary Perspective," *Review of General Psychology* 8.2 (2004): 100-110.

[76] Frans de Waal, "Putting the Altruism Back in Altruism: The Evolution of Empathy," *Annual Review of Psychology* 59 (2008): 280.

[77] Robert Trivers, "The Evolution of Reciprocal Altruism," *The Quarterly Review of Biology* 46 (1971): 35-47; Joseph Henrich, et al., "Costly Punishment Across Human Societies," *Science* 312.5781 (2006): 1767-1770.

[78] Alan Sanfey, et al., "The Neural Basis of Economic Decision-making in the Ultimatum Game," *Science* 300 (2003): 1755-1758; Jesse Graham, et al., "Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism," *Advances in Experimental Social Psychology* 47 (2013): 55-130.

These socio-moral abilities and tendencies—which evolved to facilitate intra- and intergroup cooperation and competition, as our ancestors learned to survive in diverse environments by relying on accumulating cultural knowledge and increasingly complex social dynamics—are generally understood as the building blocks of moral cognition. And yet, this evolutionary account is incomplete. It seems clear that these social dynamics structure much of our moral life. But morality is also rooted in features that are not specific to social animals. We make use of rational cognition and working memory in performing moral computations, features that many non-social animals also have. A sense of possessiveness over objects and land is also common to many non-social animals—many animals that live mostly solitary lives are prepared to defend a territory, for instance—and possessiveness and territoriality play crucial roles in our moral cognition. An ability to form attachments, to care not only about humans but also about non-human creatures, objects, and places, is also crucial to moral cognition. Finally, I will argue that an appreciation for beauty is also profoundly important for morality, and aesthetic preferences are common in many animals that we don't think of as social or moral—amphibians providing many striking examples, for instance.

There is thus a complex story of the evolution of morality that gets neglected by those who define human morality as a sub-species of human sociality. In chapter 3, I will define morality precisely and show just how important the non-social conditions of morality are. While it would be intractable to try to give a full evolutionary account of these conditions, I will focus in this chapter on troubling the conventional story of the evolution of morality by zooming in on some features of moral aesthetics that force a nuancing of current understandings. Thus, I'm going to interpret one of Dostoevsky's strangest character portraits from an evolutionary lens,

suggesting implications of this image of moral beauty for evolutionary interpretations of moral psychology.

Prelude to an Evolutionary Interpretation of The Idiot

Few have explored what it means for a person to be good, pure, or morally beautiful more thoroughly than Dostoevsky, and just like his exploration of Raskolnikov's moral depths, his portrayal of our moral heights is full of surprises and contradictions. One of his boldest character portraits, the depiction of the saintly "idiot," prince Lev Nikolaevich Myshkin, is largely a meditation on the excessive and unaccountable nature of moral beauty, a problem that provokes a nuancing of the evolutionary account sketched above, and I will read his novel *The Idiot* with an evolutionary question and sub-question in view. The question is: how might someone with such quixotic moral features have evolved? The sub-question is perhaps more to the point: how might Dostoevsky and his readers have evolved a *preference* for prince Myshkin's kind of moral beauty? What are the implications of such moral-aesthetic preferences for evolutionary interpretations of moral psychology?

With the extremity of his Christlike[79] compassion, humility, and magnanimity, prince Myshkin appears positively maladapted to many human ways of life. And yet he embodies Dostoevsky's idea of moral beauty, and some of the very qualities that render him less fit for survival simultaneously serve to captivate the hearts of the novel's most desirable young women. What gives? Why—Dostoevsky presses us with this work—why does moral beauty seem to be

[79] When writing to his niece about his idea for writing a novel portraying a "positively beautiful man," Dostoevsky draws a direct connection between the idea of Myshkin and that of Christ: "There is only one perfectly beautiful person—Christ" (Dostoevsky, *The Idiot*, from translator's introduction, p. xi).

so excessive and costly? Haven't we learned that good character is useful, that it raises one's status and greases the wheels of cooperation?

A Hans Holbein painting of Christ, dead and entombed, features emblematically in the novel, fascinating both Myshkin and the villain Parfyon Rogozhin, just as it had Dostoevsky himself.[80] This depiction is so utterly and pathetically human, so bereft of any spark of divine transcendence, that from looking at it, says Myshkin, a "man could even lose his faith."[81] "Lose it he does," Rogozhin agrees.[82] Myshkin is like this painting—a profoundly non-transcendent image of Christ—and this is what makes him so interesting to study from an evolutionary biological perspective. This representation brings the theological down into the dirty messy world of nature and shows how the two relate. With the character of Myshkin, Dostoevsky poses some tough questions, not only for Christian theologians but also for evolutionary theorists. Can our theories accommodate *that*?

As the title of the work suggests, Myshkin's moral "perfection" has a strained relation to his intellect. It's not that the prince isn't smart. On the contrary, while others initially view the prince's generosity of spirit as "idiotic," they are soon astounded by his perspicacity: he seems to see right into the souls of other people—to perceive their real motives, even when they dissimulate—and this makes his graciousness all the more surprising. It also puts Myshkin in a difficult psychological position. His rational faculty is in dynamic tension with his humility and charity, and the result is a dangerous internal instability, as he struggles to maintain subjective coherence.

---

[80] Dostoevsky, *The Idiot*, translator's introduction, pg. ix.

[81] Dostoevsky, *The Idiot*, 218.

[82] Ibid.

As with Raskolnikov, the struggle to maintain coherence in the face of these internal contradictions contributes to physical illness. The excessive nature of Myshkin's Christlike "perfection" comes at a cost. One illustration of this occurs in the second of the novel's four parts, as Myshkin returns home after a meeting with his rival and friend, Parfyon Rogozhin. Myshkin is wracked with guilt for suspecting Rogozhin of having violent intentions towards Rogohzin's would-be bride, Nastasya Filippovna, and the prince himself, even though this suspicion is extremely well founded. Trying to reassure himself that Rogozhin really has "an immense heart"[83] and reproaching himself bitterly for his unkind suspicion, Myshkin begins to feel the signs of illness coming on. And then the prince is suddenly confronted with Rogozhin himself—eyes flashing, knife raised. "Parfyon, I don't believe it!" the prince cries out, and falls into an epileptic fit.[84]

From this and many other scenes in *The Idiot*, it becomes clear that moral beauty, as Dostoevsky understands it, can be excessive and costly, dangerous to its bearer and even to others. I take this as a provocation for evolutionary accounts. Accepting the intuitive sensibleness of the sketch of the evolution of our moral-psychological qualities with which I began this chapter, I also acknowledge its insufficiency for accommodating the "higher realism" of Dostoevsky's meditation on moral beauty. Below, I elaborate on this problem and point towards a solution.

Evolution & the Riddle of Moral Beauty

[83] Fyodor Dostoevsky, *The Idiot*, trans. Richard Pevear & Larissa Volokhonsky (New York: Vintage Books, 2003), 230.

[84] Ibid. 234.

Natural Selection, Sexual Selection

With the publication of *On the Origin of Species,* in 1859*,* Charles Darwin offered a plausible, empirically grounded theory for how life could evolve from simple forms into a diversity of organisms whose features are shaped by an ongoing struggle to survive and reproduce within competitive ecological systems. Drawing on Thomas Malthus's theory of population growth, Darwin saw that the exponential rates at which biological species reproduce must initiate a struggle for survival, and that over time this struggle would "select" for traits that gave some organisms an advantage over others. This process of selection would proceed analogously to the way in which breeders select for traits in their animals, and over time this action of "natural selection" would create an adaptive fit between organisms and environments, a fit that would not be static but creative, leading to increasing complexity of biological forms and relations among forms. New specialized features, new antagonistic and symbiotic relations among species, even entirely new species would arise given this basic principle and enough time.

Darwin's *Origin of Species* provided an apt response to an objection that many people made to the very idea of evolution: that the evident perfection of the design fit between organisms and their environments could never have arisen through the mindless dynamics of nature but must be the work of an "intelligent" designer. With the theory of natural selection, Darwin showed how such a fit could arise without the need to posit intelligent design. And yet, by the time he published *Origin,* Darwin had an altogether different theoretical worry. All around, he saw organisms with evidently *maladaptive* traits. Could his theory of "natural selection" explain the evolution of loud calls and showy displays that render numerous invertebrates and members of every major vertebrate class vulnerable to predation? Perhaps the most offensive features are found among birds. A large, brightly colored tail makes the peacock

both slower and more conspicuous than it would otherwise be. Under natural selection, peacocks with large gaudy tails should be killed and eaten by predators more frequently than their more modest peers, and the selective pressure supplied by predation should ensure that these feathers never evolve in the first place. Thus, with such features, Darwin had to contend with a diametrically opposite problem than that raised by the intelligent designers: many features appeared to be designed altogether *un*intelligently.

In *Origin of Species,* Darwin proposed a mechanism to explain this problem of evidently unintelligent design, although he didn't explore this hypothesis in detail until the publication of *The Descent of Man and Selection in Relation to Sex,* in 1871. As males compete among each other for mates, Darwin hypothesized, they might evolve both offensive and defensive weapons to assist in battle; and as females make choices about which males to mate with, their preferences for certain kinds of adornment, coloration, or song could exert its own selective pressure.[85] Peacocks grow such beautiful tails, Darwin decided, because peahens like them, and the high mating success of well-endowed peacocks compensates for the loss in fitness they incur under natural selection.[86] The size and gaudiness of each individual peacock's tail thus expresses a compromise between competing selection pressures in the environments of their ancestors.[87]

---

[85] Male mate choice is also important in many species, including humans, and I feel we still know little about differences in sexual selection between men and women. I have chosen to forgo this discussion.

[86] Charles Darwin, *On the Origin of Species by Means of Natural Selection,* American Edition (New York: D. Appleton & Company, 1861) chpt 4, 84-85; Charles Darwin, *The Descent of Man and Selection in Relation to Sex* (London: John Murray, 1871) Volume II, p. 123-124.

[87] The real dynamics of sexual selection and natural selection are complex and do not fit neatly into a binary opposition, with adaptations on one side and maladaptations on the other. Natural selection *can* lead to imperfect adaptations, as well as non-adaptive side effects of adaptations, sometimes called "spandrels," and sexual selection very often favors adaptive traits or traits that,

Nevertheless, this hypothesis of "sexual selection" raised as many questions as it answered, and four months after the publication of *Origin,* Darwin wrote to the botanist Asa Gray that "[t]he sight of a feather in a peacock's tail, whenever I gaze at it, makes me sick!"[88]

Chief among the questions raised by the hypothesis of sexual selection is—in the first place—why? Even if we grant that the peacock's feathers have resulted from the preferences of peahens, why should peahens be attracted by maladaptive features? If conservative feathers render a peacock "fitter," in evolutionary terms, shouldn't peahens evolve to prefer more modest males whose offspring will inherit their drab coloring along with greater odds of surviving? Shouldn't sexual selection simply *reinforce* the pressure of natural selection, as members of each sex mate preferentially with the fittest members of the other sex? While sexual selection is a plausible mechanism and would ultimately prove to be an important part of the solution to the problem of "unintelligent design," Darwin did not have an explanation for why such "unintelligent" preferences should arise in the first place. Thus, his explanation simply shifted the theoretical burden to a deeper problem.

An early model that went beyond Darwin's observation about the significance of sexual choice is found in Ronald Fisher's 1930 classic, *The Genetical Theory of Natural Selection*. In this work, Fisher proposes that sexual preferences for a variety of features are likely to arise due to the normal action of natural selection. It is of obvious value, for instance, to be able to reliably

---

while not initially adaptive, are later "exapted" for some new use. For an introduction to these concepts and the structuralist evolutionary approach, see Stephen J. Gould and Richard C. Lewontin, "The Spandrels of San Marco," *Proceedings of the Royal Society of London*, 205.1161 (1979): 581-598; Stephen J. Gould and Elisabeth S. Vrba, "Exaptation—A Missing Term in the Science of Form," *Paleobiology* 8.1 (1982): 4-15.

[88] Francis Darwin ed. *The Life and Letters of Charles Darwin, Including an Autobiographical Chapter* (London: John Murray, 1887), Volume II, 296.

recognize members of one's own species, since, if you mate with a different species, your offspring are likely to be either infertile or greatly disadvantaged in the struggle to survive and reproduce.[89] Beyond this, in species where females choose among a variety of male suitors, it would be valuable for these females to be able to identify genetically well-endowed males and preferentially mate with them, since this would give their offspring a better chance of survival. A preference for traits that reliably signify high fitness could initially spread due to the action of natural selection. This is rather obvious, however. Fisher's theoretical innovation was to point out that the preference and the trait would then be linked in offspring. As females with the preference mated with males with the trait, not only would the male offspring inherit the trait, the female offspring would inherit the preference. If the strength of sexual selection were sufficiently strong, this would initiate a "runaway" process whereby both the trait and the preference for the trait would spread rapidly throughout the population.

In and of itself, such a "runaway" process does not explain how a trait as exaggerated as the peacock's tail might evolve. It only specifies the conditions under which the trait and a preference for the trait could rapidly spread throughout a population. In order to understand how the exaggerated traits that so distressed Darwin might have evolved, it is necessary to account for the fact that some sexual preferences have an inherent *directionality*: that is, that the preference that spreads due to runaway selection is sometimes for *longer* feathers, *brighter* colors, *more complex* markings, etc., rather than for something more modest. Sexual preferences needn't be directional in this way, and in fact preferences are often "stabilizing," as partners select for a

[89] Ronald Fisher, *The Genetical Theory of Natural Selection* (London: Oxford University Press, 1930), p. 130.

mean between extremes.[90] Sometimes, however, sexual preferences have a strong directional slant, and when this is the case, Fisherian "runaway" selection can lead to the remarkable exaggerations of form and behavior that so troubled Darwin.

Again, the explanatory burden was pushed back. The conditions under which sexual preferences take on a directional slant still remained to be explained. According to Fisher's hypothesis, the development of preferences for adaptive features will tend to reinforce, however slightly, the preexisting pressure of natural selection for these traits. Greatly exaggerated traits, on the other hand, are "costly," whether because they require nutritive and immune resources or because they make the organism more conspicuous and less able to evade predators. Such traits violate the balance that is maintained under natural selection, and they can only come into existence if another selective force drives their development out of all proportion. Thus, the question remains: why shouldn't peahens maintain a preference for a moderate feature that would signal the fitness of the peacock without becoming so exaggerated as to actually *reduce* the bird's fitness in so doing?

Today we are aware of at least two possibilities. For one, there is the hypothesis of "sensory exploitation."[91] Peahens might have had a preexisting bias favoring greater sensory stimulation of a certain kind. This sensory bias might have nothing to do with the fitness of the male, but might have arisen due to some entirely different process. Nevertheless, as peahens exerted sexual choice, they might effectively begin to breed males in accordance with their

---

[90] Michael Ryan, "Sexual Selection, Sensory Systems, and Sensory Exploitation," *Oxford Surveys in Evolutionary Biology,* eds. Douglas Futuyama & Janis Antonovics (London: Oxford University Press: 1990) 7: 157–195.

[91] Ryan, "Sexual Selection, Sensory Systems, and Sensory Exploitation," 157-195; Michael Ryan, et al., "Sexual Selection for Sensory Exploitation in the Frog Physalaemus Pustulosus," Nature 343 (1990): 66-67.

misguided taste. And as a "runaway" process began and the strength of sexual selection increased, the feature would become increasingly exaggerated, until it either fully satisfied the peahens' preexisting predilection or was overwhelmed by opposing pressure from natural selection. The "final" version of the peacock's tail (the point at which its evolution reached relative stasis) would express a compromise between the costs of the trait under natural selection and the benefits accruing to peacocks whose features exploited this sensory bias of peahens.

In 1975, Amotz Zahavi proposed a different, strikingly counterintuitive solution to the riddle of the peacock's tail. According to his "handicap" hypothesis, the directionality of the peahen's preference may have developed precisely *because* of the tail's fitness costs. As Zahavi pointed out, the very fact that the tail reduces fitness (and in this sense functions as a "handicap") means that only the relatively fitter males can afford to grow such tails and live to show it. The unwieldy tail thereby functions as an "honest" signal of the exceptional fitness of its bearer. Merely surviving with such an extravagant tail is proof of the peacock's exceptional fitness. He has passed a kind of "test," and as long as the reliability of the information gained through this test is sufficiently valuable to peahens making a mating choice, the handicap will be selected for, even in the face of the substantial fitness costs of an extravagant tail. Under this hypothesis, the "final" version of the tail expresses a different compromise: this time between the initial costs of the trait under natural selection and the benefits of the *information* the peahen gains from the trait, which she uses in making her mate choice.[92]

These solutions have very different implications. You might say that "sensory exploitation" really does lead to rather "unintelligent" design features. The traits that arise under

---

[92] Amotz Zahavi, "Mate Selection – A Selection for a Handicap," *Journal of Theoretical Biology* 53 (1975): 205–214.

this kind of selective pressure may be totally maladaptive, but if pressure from sexual selection is strong enough, the traits will arise anyway. By contrast, a preference for a "handicap" remains connected to the underlying value that the trait signals, even if a "runaway" process is initiated. By hypothesis, in this latter case the trait initially arises because it carries valuable information about the fitness of its bearer. As the trait subsequently becomes exaggerated, the information that it provides becomes increasingly reliable—proof that its bearer has passed a "test." Thus, the peahen with a preference for a "handicap" has a superior sensitivity to quality, a sensitivity that is ultimately quite useful.[93]

The Evolution of "Altruism"

Recall our formulation of the problem of "unintelligent design": if evolution occurs through selection for features that increase fitness and selection against features that decrease fitness, how can a fitness-*decreasing* feature ever evolve? In many cases, as we have seen, the answer to this riddle is that certain fitness-decreasing features can evolve through sexual selection—whether because they exploit a preexisting sensory bias of the opposite sex, or because they function as an "honest" signal of fitness, proof to potential mates that the bearer has passed a "test." But there is another category of such fitness-decreasing features that has been obsessed over by evolutionary theorists for pretty much ever, and that has a more obvious connection to our topic. The problem of "altruism" is the problem of explaining the evolution of features that decrease the fitness of their bearer *while increasing the fitness of someone else*. There is some major overlap between this riddle of the evolution of "altruism" and that of the evolution of sexy

[93] Even though these hypotheses are theoretically distinct, I see no reason why they may not both exert influence upon the evolution of a trait. In such cases, it would be more difficult to reach conclusions as to the value of the trait and the preference.

features, but it is worth considering "altruism" in its own right, not only for theoretical reasons but also for historical reasons: strikingly many of the most important theoretical innovations in evolutionary theory have arisen in response to the riddle of "altruism," and solving this riddle has been seen by many as key to understanding the evolution of morality.[94]

Solutions to the "altruism" riddle include William Hamilton's model of inclusive fitness, or "kin selection,"[95] Robert Trivers's "reciprocal altruism,"[96] Amotz Zahavi's "handicap" theory of sexual selection for altruism,[97] Martin Nowak's "indirect altruism,"[98] recent "cultural group selection" models of human cooperation, [99] and biological group selection models, [100] including Herbert Gintis's and others' "strong reciprocity."[101] As this proliferation of models suggests, there are actually a variety of solutions to this problem. "Altruism," as defined above, includes a wide range of evolved behaviors and physical features, and even a single behavior typically

---

[94] They are wrong. Altruism is not even close to the whole story of morality.

[95] William Hamilton, "The Genetical Theory of Social Behaviour," *Journal of Theoretical Biology* 7 (1964): 1-52.

[96] Robert Trivers, "The Evolution of Reciprocal Altruism," *The Quarterly Review of Biology* 46 (1971): 35-47.

[97] Amotz Zahavi, "Altruism as a Handicap: the Limitations of Kin Selection and Reciprocity," *Journal of Avian Biology* 26 (1995): 1–3.

[98] Martin Nowak and Karl Sigmund, "Evolution of Indirect Reciprocity," *Nature* 437.27 (2005): 1291-1298.

[99] Richerson, Peter & Robert Boyd, *Not By Genes Alone: How Culture Transformed Human Evolution* (Chicago: University of Chicago Press, 2005); Henrich & Henrich, *Why Humans Cooperate*, 2007.

[100] Elliot Sober and David Sloan Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Cambridge: Harvard University Press, 1998).

[101] Gintis et al., "Strong Reciprocity and the Roots of Human Morality," 241-253.

arises through a push and pull among a plurality of selective forces. Thus, even as certain "altruism" riddles have been solved, this should by no means be taken to imply that theorists have accounted for the evolution of all "altruistic" features.

In the first part of this chapter, I drew upon several of the above models to help characterize the evolution of human cooperation. Beyond this, an important lesson we should learn from the history of responses to the "altruism" riddle is an understanding of how altruism *did not* evolve. One flawed solution to the "altruism" riddle was offered by early proponents of "group selection," such as the fantastically named Vero Copner Wynne-Edwards, who believed that prosocial behaviors must be favored by selection at the level of *groups* of organisms, and that selection at this level would, as a general rule, win out over selection for selfishness at the individual level.[102] George Williams famously demonstrated the error of this conception of group selection in *Adaptation and Natural Selection* (1966), offering both empirical and theoretical arguments to show how selection at the individual level could better explain many of the behaviors that Wynne-Edwards and others interpreted as group-level adaptations. For example, whereas Wynne-Edwards maintained that intrasexual competition for mating rights was a group-level adaptation, designed to limit population size and prevent overcrowding, Williams pointed out that under this assumption, we wouldn't expect *males* to do the competing, since, even though such competition reduces the pool of males that get to breed in a given season, the males that win these competitions tend to mate with most of the females, and population size is little affected.[103] Again and again, Williams showed that phenomena that we observe in groups of

---

[102] Vero Copner Wynne-Edwards. *Animal Dispersion in Relation to Social Behavior* (Edinburgh: Oliver and Boyd, 1962), 20.

[103] George Williams. *Adaptation and Natural Selection* (NJ: Princeton University Press, 1966), 245-246.

animals can be explained simply and parsimoniously by invoking selection at the individual level. A "fleet" herd of deer is not a herd that has been designed by group selection to outrun other herds but rather a herd of individual deer that have each been designed by competition with the other members of their group.[104] New theories of group-selection that attempt to respond satisfactorily to Williams's critique have been published since, notably in the work of Elliott Sober, David Sloan Wilson, and E.O. Wilson, and today there is an ongoing debate about whether group-selection has played an important role in the evolution of "altruism" in humans. However, all parties to this debate agree in their rejection of the naïve group-selectionism that enjoyed some popularity prior to the work of George Williams.

Another more persistent error is the belief of some biologists that their models are intrinsically "designed to *take the altruism out* of altruism"[105]: i.e., to show how acts of seeming kindness and self-sacrifice are really only a kind of veneer over a fundamental reality of individual selfishness.[106] In order to address this error, we must begin with an appreciation of how different the biological definition of "altruism" is from our colloquial understanding of the term, the latter which essentially involves an imputation of altruistic *motives*. As Richard Dawkins has argued, a gene favoring tooth decay in lions would be a gene for "altruism," under the biological definition above, if it slowed down the lion's rate of meat consumption and

---

[104] Williams, *Adaptation and Natural Selection,* 16-17.

[105] Trivers, "Reciprocal Altruism," 35, emphasis mine.

[106] Frans de Waal, et al. *Primates and Philosophers,* (New Jersey: Princeton University Press, 2006).

allowed other members of her pride (including close relatives) to eat more.[107] But obviously, such a gene would not promote altruism, as the term is normally understood. The lioness with the bad teeth isn't restraining herself out of largesse, but rather failing—due to a structural deficiency—in her attempt to eat as much as she possibly can.

Such specialized language-conventions are useful, as long as they are used consistently, but we must be on guard against equivocation. When these same biologists turn to the implications of their models, they sometimes fail to maintain the separation of the biological definition of "altruism" from its colloquial meaning—thereby failing to maintain an important theoretical separation of the "ultimate" evolutionary level of analysis from "proximate" considerations of motivation and function.[108] The suggestion that evolutionary explanations of "altruism" are intrinsically designed to undermine evidence for the importance of compassionate and generous motives is plainly wrong. While "altruistic" evolutionary outcomes among humans may be favored by a variety of proximate motives, it is well established that empathy, for instance, is one such motive.[109] And the mere fact that in many cases individuals stand to gain something in return for their altruistic actions does not mean that their acts are selfishly motivated. Magnanimous deeds are not always motivated by empathy, and they may involve gains to the magnanimous individual (on average, and in the long run), but this in no way carries

---

[107] Richard Dawkins, "Twelve Misunderstandings of Kin Selection" *Zeitschrift für Tierpsychologie* 51 (1979): 190.

[108] Ernst Mayr, "Cause and Effect in Biology" *Science* 134.3489 (1961): 1501-1506; Sober, Elliot and David Sloan Wilson. *Unto Others*, chapter 2.

[109] Frans de Waal, "Putting the Altruism Back in Altruism: The Evolution of Empathy," *Annual Review of Psychology* 59 (2008): 280.

the implication that bravery, loyalty, and generosity are mere hypocrisy, a veneer over a reality of selfish ambition.

Moral Beauty as a Handicap

Our use of evolutionary models of "altruism" has so far focused on explaining the evolution of a human psyche that is predisposed towards sensible, adaptive forms of cooperation, but the provocation Dostoevsky makes with the character of prince Myshkin is altogether different. Why should moral beauty be *mal*adaptive? Why should we have evolved a preference for traits that impose such a stark fitness cost? That is, why should our preference for morally beautiful traits be *directional* rather than stabilizing? As opposed to the "altruism" promoted by kin selection or reciprocal altruism, this remains an unanswered provocation, analogous to the problem of "unintelligent design" posed by the peacock's tail.

Conveniently enough, twenty years after proposing the "handicap" hypothesis of sexual selection for striking plumage, Zahavi suggested that the same hypothesis could also explain certain forms of "altruistic" behavior.[110] Recall that a reduction in the fitness of the altruist is part of the evolutionary definition of "altruism." Zahavi argued that, since "altruistic" behavior reduces the fitness of the altruist, such behavior could also function as a "handicap," similar to the "handicapping" effect of extravagant plumage. Individuals who could afford to be "altruistic" would tend to have better genetic or material resources than selfish individuals, and "altruism" could thus function as a reliable signal of fitness—proof that the altruist had passed an

[110] Zahavi, "Altruism as a Handicap," 1–3. Why don't humans have extravagant physical features like many other species? Zahavi notes that such features are more common among species that pair bond quickly, without much time to assess quality. Humans have a variety of ways to assess quality, and a good deal of time to do so. Thus, we may not need to develop physical "handicaps." Spiritual ones will do just fine.

evolutionary "test." From this, a sexual preference for more "altruistic" individuals could evolve, as selective pressure favored the very trait that was being selected against under natural selection (with the informational value of the "honest" signal compensating for the cost of the trait under natural selection).[111]

As readers of Dostoevsky's novel, we have access to some of Myshkin's internal motivations: his sense of compassion, which overrules his critical judgments (despite his remarkable perspicacity), his courage, his practice of putting others before himself, even in thought. If we admire Myshkin, it is not because these qualities render him fitter under natural selection at the individual level—most evidence is to the contrary—nor is it because his thought and behavior adhere to a coherent set of moral laws—his charity and humility lead him to violate other moral norms relating to rational utility, fairness, and justice—nor even is it because his virtue increases the fitness of his group—it clearly doesn't. In spite of such "idiocy," virtually all who come to know Myshkin come to admire him. Thus, just as Dostoevsky's depiction of Myshkin contradicts rationalistic philosophical ideas of moral virtue, it also challenges straightforward adaptationist accounts that would purport to compass the evolution of moral beauty. What makes a person's character beautiful? Is it their ability to always find the perfect mean between unethical extremes? Is it their consistent subordination of their desires to the dictates of the moral law, as determined by reason? Is it their rigorous adherence to the principle of maximal utility? Is it their ability to forge valuable cooperative relationships? Or is moral

[111] Zahavi actually makes the second error described above when it comes to interpreting his model, claiming that the "handicap" model purports to explain "altruistic" behavior as being selfishly motivated. It is important to maintain the separation between proximate and ultimate levels of analysis, however. Much "altruistic" behavior certainly is motivated selfishly, but certainly not all. Even when one's altruism is advertised and leads to greater sexual access, it may be misleading to say that it is selfishly motivated.

beauty something excessive and strange, like the "perfection" Jesus called his followers to embody? A signal of quality that is authenticated by its high cost; a sign that its bearer has passed a "test"?[112]

I will answer these questions with a kind of parable. The main point of this parable, I'll just say it, is that moral beauty—just like other forms of beauty—is not a one-dimensional thing. It's complicated. Some kinds of virtue can be explained by the evolutionary story with which I began the chapter, but this explanation falls far short of accounting for other kinds of moral beauty. To get the full picture of the evolution of moral cognition, we must confront the things that don't immediately make sense. Just as Darwin was troubled by the sight of a peacock's feather, we should be troubled by Dostoevsky's portrayal of moral beauty, realizing the challenge it poses for evolutionary theory.

Moral Beauty in The Idiot

Aglaya Ivanovna Epanchin is an extraordinarily beautiful young woman,[113] just reaching marriageable age, of good birth and with the promise of a large dowry. As she considers an array of suitors, Aglaya has every reason to be choosy in selecting "the one," and every reason to be skeptical about their professions of love. Thus, she works hard to perceive the true motives and character of her suitors, and she is ready to put them to shame if they so deserve. The very first

[112] In spite of my provocation, I don't really think we have to choose just one answer here. There are distinct dimensions in which we judge beauty (consider that the peacock is not only rendered beautiful by his excessive plumage but also by more modest signals of fitness, selected by stabilizing preferences for facial symmetry, BMI, etc.).

[113] When asked about Aglaya's beauty, prince Myshkin claims he is not ready to interpret it, remarking that "[b]eauty is a riddle" (Dostoevsky, *The Idiot*, 77).

time we meet Aglaya, we observe her doing precisely this to a young suitor who has failed to meet her standards.

Gavrila Ardalionovich Ivolgin (aka. Ganya) is prideful and incredibly ambitious. But lately he has been mortified by his family's fall from good society, as a result of his father's buffoonery, open marital indiscretions, and failure to honor his many debts. Determined to leave behind this disgrace by growing rich and clawing his way back up the social hierarchy, Ganya has become embroiled in multiple marriage plots. For one, he has been offered a large sum of money to marry Nastasya Filippovna, a young woman whose great beauty[114] strikes prince Myshkin as "dazzling," "strange," and "unbearable."[115] Nastasya has recently emerged into Petersburg society after having been orphaned in childhood and then kept during her teenage years as the personal prostitute of her adoptive benefactor, a high-society man named Afanasy Ivanovich Totsky. Now in his mid-fifties, Totsky is finally ready to get married—not to Nastasya, since her history as a "kept" woman (*his* kept woman!) makes her unacceptable—but to a young woman who would be deemed appropriate for someone of his social position (he has designs on Aglaya's older sister). But Nastasya has been threatening to make a scene, and Totsky figures that the only way to protect himself from some embarrassment is to marry her off—hence his offer to Ganya. Totsky and General Ivan Fyodorovich Epanchin (Aglaya's father) have approached Nastasya Filippovna with their proposal, as it were on Ganya's behalf, and she has promised to give her answer at her upcoming birthday party.

---

[114] "You can overturn the world with such beauty," says Aglaya's sister—an artist—upon viewing a portrait of Nastasya (Ibid., 80).

[115] Dostoevsky, *Idiot,* 79-80.

Even as Ganya is participating in this plot, he is holding out hope for an even better prospect. A marriage to Aglaya would bring not only a beautiful woman with a large dowry but also immediate social connections. Prince Myshkin meets Ganya just at the moment when the latter must decide, once and for all, whether to pursue Nastasya or Aglaya. Characteristically, however, as someone who is primarily driven by self-interest, Ganya tries to hedge his bets. He asks Myshkin to deliver a letter to Aglaya, a letter in which he promises not to go through with the engagement to Nastasya if Aglaya will give him just one "word of compassion."[116]

Aglaya understands much by this letter. She has observed Ganya carefully, and she is quick to recognize that he is trying to make a deal for her. She understands fully that the love he professes for her is far weaker than his selfish ambition, and she is insulted by the brazenness of his plotting. With the letter, Ganya has thus failed an important test, and Aglaya's response is vindictively dismissive: "I don't negotiate."[117] But she doesn't stop there. She also sees fit to punish him through public humiliation. Returning Ganya's letter to prince Myshkin, Aglaya tells her new male acquaintance to read it aloud. Then, she offers her assessment to the prince:

> "This man assures me," Aglaya said sharply, when the prince had finished reading, "that the words *break it all off* will not compromise me or commit me in any way, and, as you see, he gives me a written guarantee of it by this very note. See how naively he hastened to underline certain words and how crudely his secret thought shows through. He knows, however, that if he broke it all off, but by himself, alone, not waiting for a word from me, and even not telling me about it, without any hope in me, I would then change my feelings for him and would probably become his friend. He knows that for certain! But his soul is dirty…"[118]

---

[116] Dostoevsky, *Idiot,* 84.

[117] Ibid., 83.

[118] Ibid., 84

Aglaya is able to read into Ganya's character quite easily, from rather subtle signs. Her perception that Ganya is bargaining for her is not made on the basis of any explicit admission of this fact on his part—quite the contrary. And yet, she easily connects the dots. If he weren't trying to manipulate Aglaya, he would break off the relationship with Nastasya of his own free will, "without any hope" in a match with Aglaya, and the fact that he has not already done so is evidence of the impurity of his motives: "his soul is dirty."

I note that Ganya is handsome, capable, and ambitious, but for Aglaya this is not enough. It matters very much whether he has a pure "soul." Perhaps Ganya will not be a faithful husband, will not prioritize the interests of his wife and children when doing so would inconvenience him, will not provide for them and defend them when they need it the most. More to the point, perhaps he is not worthy of her respect or capable of inspiring her love. Aglaya shows disdain for Ganya's pragmatism, his desire for money and an advantageous marriage—desires that are perfectly natural, given what we know about evolution. Aglaya's disdain should give us pause.

As it turns out, Nastasya is even more offended by Ganya's and Totsky's scheming than Aglaya is. And she devises a series of dramatic tests, designed to provoke and shame them, at her birthday party, in the presence of a crowd that includes prince Myshkin, Parfyon Rogohzin (another of Nastasya's suitors, the son of a wealthy capitalist), and Ganya's co-conspirators—Totsky and General Epanchin.

Totsky has promised Ganya a dowry of seventy-five thousand roubles (equivalent to millions of dollars today) to marry Nastasya, and Nastasya is deeply offended by this. What would it say about *her* soul, if she could be so easily bought? At the same time, Rogohzin has arrived on the scene with the promise of a massive inheritance in his near future and brazenly offered Nastasya cash for her hand in marriage (a tactic he would only consider given Nastasya's

unsavory social "position"). Nastasya tells Rogohzin to arrive at her party at the time that has been set for the announcement of her engagement to Ganya, and to bring with him a hundred thousand roubles.

The many dramatic events of this party all ensue from Nastasya's determination to reveal the true character of the men who are bargaining for her. Hell-bent on exposing everyone's soul, Nastasya curates a series of dramatic confrontations. First, she initiates a parlor game, in which each of the guests is asked to confess the "worst thing" they have ever done. This is above all a provocation to Totsky, an invitation for him to do public penance for his sins. Totsky is aware of the harm he has done to Nastasya's moral reputation and the way in which she has internalized the idea that she is dirty and morally base. But he says nothing of his abuses, and in response to Nastasya's provocation he tells a cleverly self-serving anecdote, a story that shows off his connections to royalty and confesses to only a certain morally ambiguous mischievousness of his youth.

Totsky's unapologetic confession insults and infuriates Nastasya, and she responds by escalating her game. When it comes time for the promised engagement announcement, Nastasya shocks everyone by turning to prince Myshkin—a man she met earlier that day—and asking him to decide for her. At Myshkin's word, Nastasya rejects the proposal and the seventy-five thousand. In response to the protests of General Epanchin, she explains her action thusly: "The prince is this for me, that I believe in him as the first truly devoted man in my whole life. He believed in me from the first glance, and I trust him."[119] Just as Ganya has failed his character test, Myshkin has passed his. In very little time, Nastasya has perceived the prince's soul and become convinced of the accuracy of her impression.

---

[119] Ibid., 153-155.

Just at this moment, Rogohzin arrives with the hundred thousand and a rowdy entourage. Nastasya publicly considers her options: take the hundred thousand and accept that she is "Rogozhin's kind of woman," or give up everything and become "a washer-woman."[120] Considering the latter option, she asks, "who will take me without anything?" And to this rhetorical question, she receives a response from the local fool, her outspoken acquaintance Ferdyshchenko: "Maybe Ferdyshchenko won't take you…but the prince will!"[121] Everyone, including Ferdyshchenko, is busily engaged in observing each other, and from his observation of the prince,[122] Ferdyshchenko has reached this conclusion regarding Myshkin's feelings, intentions, and character.

Intrigued by this claim, Nastasya publicly forwards the question to Myshkin: "You'll take me just as I am, with nothing?"[123] At his affirmative response, Nastasya wonders openly about Myshkin's mental health: is it true that he's "*like that?*"[124] But Myshkin's subsequent explanation indicates something altogether different. Nastasya has demonstrated the ascendancy of her moral ideals over pragmatic considerations, rejecting Totsky's bribe at great personal cost. Nastasya herself has passed a test, and he has fallen in love with the beauty of her soul:

> You've given Mr. Totsky back his seventy thousand and say you will abandon everything you have here, which no one else would do. I…love you…Nastasya Filippovna. I will die for you, Nastasya Filippovna. I won't let anyone say a bad

---

[120] Ibid., 161-163.

[121] Ibid., 163.

[122] Ibid. "I've been watching him for a long time."

[123] Ibid.

[124] Ibid.

> word about you, Nastasya Filippovna…If we're poor, I'll work, Nastasya Filippovna.[125]

Before Nastasya can respond to these words, the whole ensemble is shocked at an unexpected revelation: the penniless prince is himself about to come into a substantial fortune.[126] Here, altogether unexpectedly, Nastasya is graced with the opportunity to start anew with a good man, a man who genuinely respects her, a man of aristocratic blood who, as it turns out, is also financially well-endowed. It is all too much, however. Nastasya has internalized the view of those around her that she is to blame for her years with Totsky, and she does not believe she is worthy of such happiness. Moreover, she is afraid of ruining the prince's reputation. Thus, after accepting Myshkin's offer and entertaining the prospect of a blissful future for a few manic minutes, Nastasya abruptly changes her mind and decides to run off with Rogozhin, offering the prince a parting explanation for her self-destructive choice:

> I dreamed for a long time, still in the country, where [Totsky] kept me for five years, completely alone […] and I kept imagining someone like you, kind, honest, good, and as *silly* as you are, who would suddenly come and say, "You're not guilty, Nastasya Filippovna, and I adore you!" And I sometimes dreamed so much that I'd go out of my mind…And then this one would come: he'd stay for two months a year, dishonor me, offend me, inflame me, debauch me, leave me—a thousand times I wanted to drown myself in the pond, *but I was base*, I had no courage—well, but now…Rogozhin, are you ready?[127]

Determined to rectify her "base" failure to drown herself, Nastasya tries a more indirect route to self-destruction. Before she leaves, however, she has one final test to administer—this time to Ganya, whose dirty dealings she has not forgotten:

---

[125] Ibid., 164.

[126] Ibid., 164-165.

[127] Ibid., 170-171. Emphases mine. Note the striking maladaptiveness of what Nastasya considers to be noble in this situation—suicide.

> Well, then listen, Ganya, I want to look at your soul for the last time […] Do you see this packet? [containing the hundred thousand] […] I'm now going to throw it into the fireplace, onto the fire, before everyone, all these witnesses! As soon as it catches fire all over, go into the fireplace, only without gloves, with your bare hands, with your sleeves rolled up, and pull the packet out of the fire! If you pull it out, it's yours […] And I'll admire your soul as you go into the fire after the money.[128]

Such a public exposure of his "soul" is too much for even Ganya, greedy as he is. He starts to walk away from the smoldering money—and faints.[129]

Conclusion

I began this chapter with a fairly straightforward explanation of the evolution of socio-moral cognition. As a highly social species, humans benefit from psychological and cultural mechanisms that facilitate efficient cooperation and give individuals and/or groups an edge in various forms of competition. However, my reading of Dostoevsky stretches this pragmatic framework to a breaking point. Actions and sentiments that are beautiful, it seems, may sometimes be irrational and destructive. What are we to make of this?

The reality is that our psyche is full of contradictions. This is the case in part because we have evolved to navigate conflicting selective pressures. Like all biological organisms, we are, in ethologist Nikolaas Tinbergen's phrase, "a compromise."[130] What I have specifically attempted to demonstrate, with the example of prince Myshkin and the discussion of sexual selection for

---

[128] Ibid., 171.

[129] Ibid., 173.

[130] Nikolaas Tinbergen, "On the Aims and Methods of Ethology," *Zeitschrift für Tierpsychologie* 20 (1963): 410–433.

altruistic features, is that in certain cases this compromise might lead to rather exaggerated and, in a qualified sense, *mal*adaptive features.

I also mentioned, without offering an evolutionary story, some other features that are important to moral cognition but that are not exclusively social adaptations: working memory, rational cognition, possessiveness, and valuation of non-human beings. I will expand upon this claim in the next chapter, showing that moral cognition does not fit within the domain of social cognition, as has typically been assumed by moral psychologists. Challenging the traditional account, in chapter 3 I trace the real complexity of our moral psychology at the proximate level of cognition.

Chapter III: Moral Cognition & Being

Chapter III Contents

Major Features of Socio-Moral Cognition

What is moral cognition? Moral psychologists can point to instances of moral cognition, but we have not figured out the necessary conditions of morality or moral cognition, and this absence of a transcendental definition has led to quite of a bit of confusion. Different researchers focus on different trees, and we all miss the forest. In this chapter, I will look at both the forest and the trees. But I will begin with the trees, critically analyzing and synthesizing existing research into moral cognition, including research on moral heuristics, theory of mind, character attribution, moral sentiments, value pluralism vs. monism, and implicit vs. explicit cognition. I conclude the chapter with a second reading of Raskolnikov's confession.

Moral Heuristics

Imagine a train bearing down the tracks towards five innocent people, who have been tied up and left to be squashed. You are close to a lever, which when pulled will transfer the train to another track, causing it to miss the five. The only problem is, there is a single person who has been tied up and left on the other track, so that if you pull the lever the train will be diverted from the five but kill the one. If you are a typical person placed in this dilemma and given the choice to pull the lever or refrain, you will pull the lever, thus killing the one to save the five. Such a decision accords with a consequentialist / utilitarian logic that says that what is right to do is what brings about the greatest good—although it seems to violate a common deontological rule to the effect that "thou shalt not kill."[131] Now imagine that there is no lever, and the only way to

[131] Exodus 20:13. Consequentialist principles involve an emphasis on the consequences of actions and a sense that "the ends justify the means" (and utilitarianism is a specific form of consequentialism that suggests we should base moral decisions on a rational calculation of the total pleasure and pain that will result). Deontological principles, in the simplest sense, are principles of right and wrong that one must follow regardless of the consequences. In

stop the train from killing the five is to push a large stranger onto the track (suspending any uncertainty about the outcome of such an action with which one would be faced in real life). Now, if you are a typical person, you will balk at such an action and choose instead to let the five die. Here, it seems, the deontological injunction against killing tends to trump the basic utilitarian dictum.

These scenarios are two of the more famous "trolley car dilemmas," moral dilemmas dreamed up by the philosopher Philippa Foot in 1967 and subsequently immortalized by teachers of Ethics 101 and, most famously, cognitive neuroscientist Joshua Greene. Through the various scenarios that have been presented to participants, psychologists have tried to probe the cognitive processes that give rise to our intuitions about what is permissible and forbidden—at least in the artificial contexts of these moral dilemmas. For instance, the pattern of responses above suggests that typical people might use both consequentialist and deontological heuristics to figure out the right course of action, with contextual factors influencing which heuristic is privileged. In this case, the difference between pulling a lever vs. pushing someone onto the tracks seems to affect which heuristic wins out.[132] Evidence that such details systematically affect moral judgment

---

deontological ethics, these principles are logically deduced from more basic principles—most famously, Immanuel Kant's "categorical imperative." In moral psychology, however, deontology is often used in the simple sense just given, which may be more appropriate for talking about the moral reasoning of non-philosophers in average everyday situations.

[132] However, there is a deontological interpretation of this finding that invokes the so-called doctrine of double effect (DDE), which basically says that you may cause a bad thing to happen if it is only a side-effect of something good that you are doing, but you may not deliberately do something evil in order to accomplish something good. Some proponents of a deontological interpretation of these findings say that pushing the stranger violates the DDE, whereas pulling the lever does not; and therefore, people's intuitions in both scenarios are consistent with deontological reasoning. There are further scenarios that try to get at finer-grained distinctions, and interpretations that challenge the deontologists' interpretation, which I will not get into.

seems to offer insight into basic features of moral cognition, a fact that has been thoroughly exploited by psychologists over the past couple decades.

For instance, using fMRI to image the brains of people as they reasoned about trolley car and similar dilemmas, Greene found that activity in brain regions associated with "emotion and social cognition," such as medial prefontal cortex (mPFC), precuneus, and the temperoparietal junction (TPJ), is associated with people's reasoning about up-close-and-personal scenarios, like the dilemma that requires people to decide whether or not to push a man to his death; whereas, activity in regions associated with "abstract reasoning and problem solving," such as dorsolateral prefrontal cortex (dlPFC), tends to be higher during deliberation about less personal dilemmas, such as the decision whether or not to pull a lever to save five people but kill one.[133] Using such fMRI and behavioral evidence, Greene has proposed a "dual process" model of moral psychology, suggesting that brain systems involved in deliberative reasoning sometimes compete against systems involved in generating strong social feelings. In this view, the different heuristics we use to arrive at intuitions about right and wrong in moral dilemmas may be said to arise from a polyphonic organization of our biology. We are endowed with a plurality of brain systems that each frame the same moral situation differently. Without any way to harmoniously resolve these framings, we are left with a zero-sum competition between neural interlocutors that express irreducibly distinct moral priorities. This Dual-Process theory is obviously compatible with my polyphonic model, and I will only add that Greene's dual-process is one source of polyphony, and it is not the primary source—but more on that later.

---

[133] Joshua Greene et al., "The Neural Bases of Cognitive Conflict and Control in Moral Judgment," *Neuron* 44 (2004): 390.

Analyzing response patterns to a variety of moral dilemmas, legal scholar John Mikhail has suggested that our ability to form rapid and predictable judgments about moral violations might develop from a specially evolved moral faculty, which he calls the Universal Moral Grammar, similar to Noam Chomsky's idea that humans have evolved a language faculty that facilitates language acquisition.[134] I will not offer a detailed analysis of Mikhail's theory except to point out that, if humans do have a faculty that helps us deliberate about moral rules, in a lawyerly fashion as Mikhail purports, this fact can be accommodated by a polyphonic model of moral psychology. But within a polyphonic model, this faculty would only be one component among several others, and hardly the most important.

Here's why. Aside from the interesting fact that there are regularities in people's intuitions about how to resolve moral dilemmas, one thing we should note about all these stories that philosophers and psychologists have used to probe people's moral intuitions, is that they are abstracted away from the kinds of relationships in which people are normally embedded, which normally determine much of how we feel about what we ought to do. These dilemmas force participants into the highly unusual position of acting as an impartial judge. The fact that we are *able* to adopt such a perspective—however uncomfortably—and that when we do so, our decisions follow predictable patterns, can only tell us so much about how we perform moral judgments in ecologically typical conditions. In everyday reality, moral sentiments and judgments arise in view of our personal involvements. For instance, if two of my friends have recently broken up, and they are both telling me what a bad person the other one is, I might find it difficult to decide who is *really* right and who is really wrong. I may simply side with whichever party I'm closest to. Such parochialism is a very different kind of heuristic than those

[134] John Mikhail, "Universal Moral Grammar," *TRENDS in Cognitive Sciences* 11.4 (2007): 143-152.

proposed by Mikhail, though I suspect that it is as "universal" as anything posited in his Universal Moral Grammar. Thus, the requirement that we take up the perspective of an impartial observer precludes a priori what is often most decisive for moral judgment in the real world.

In contrast, I feel that any model of moral psychology must accommodate the fact that in an acrimonious breakup, each party often feels him or herself to be in the right—and each person's friends typically agree with their friend—and deliberation about moral principles rarely changes these convictions. Such parochialism, it seems to me, is an image of most moral judgment, most of the time. Our personal involvements with people, places, creatures, objects, and ideas determine the very domain of morality, and strongly shape our understanding of how we ought to perceive, judge, behave, and even feel in every situation. Seeking for moral objectivity, moral psychologists have also failed to appreciate what is most decisive for moral judgment and most critical for determining the moral domain. Even psychologists who emphasize the role of irrational feelings and parochial loyalties, such as Greene and Haidt, do not get to the root of what is going on. So, I attempt to do so below, in the section on "existential framing." First, however, I will touch on a few other cognitive abilities / tendencies that psychologists have recognized as being crucial to moral judgment.

Character Attribution and Theory of Mind

In chapter 2, I mentioned that there has been a recent revival of interest in moral character. Although it is undeniable that humans keep track of moral rules and pay special attention to rule violations, some theorists argue that the primary object of moral judgment is not

the action but the person who acts.[135] What we really need to know is not whether a rule has been violated but, ultimately, whether we can trust someone in a cooperative relationship: whether they will be good spouses, neighbors, friends, and allies, or whether they are merely using us for their own ends and are ready to cheat or betray us if given the opportunity.[136]

Keeping track of others' compliance with or violation of moral rules is one way to gain such information, but there are many other ways, and lots of evidence shows that moral judgment is about much more than rules. A hypothesis that is gaining traction is that character is a central focus of moral judgment, and this person-centered approach is supported by several lines of evidence. For instance, people do not care simply about whether an action violates a moral rule or not, but are highly attuned to cues that indicate whether the violation is impulsive or coldly intentional.[137] We perceive certain kinds of violations as being especially informative about character, even when we do not view them as especially serious moral infractions.[138] We

---

[135] Erich Uhlmann, et al., "A Person-centered Approach to Moral Judgment," *Perspectives on Psychological Science* 10 (2015): 72-81; David Pizarro & David Tannenbaum, "Bringing Character Back: How the Motivation to Evaluate Character Influences Judgments of Moral Blame," In *The Social Psychology of Morality: Exploring the Causes of Good and Evil*, 91-108, eds. Mario Mikulincer & Phillip Shaver (Washington, DC: APA Press, 2011).

[136] Trustworthiness is not the only moral quality to which we are attuned, however. We want to know, for instance, whether someone is strong and principled in their values or weak and susceptible to compromise or corruption, whether they are rational or irrational, caring or indifferent, and so forth.

[137] David Pizarro, et al., "Asymmetry in Judgments of Moral Blame and Praise: The Role of Perceived Metadesires," Psychological Science 14 (2003): 267--272.

[138] Erich Uhlmann & Luke Lei Zhu, "Acts, Persons, and Intuitions: Person-centered Cues and Gut Reactions to Harmless Transgressions," *Social Psychological and Personality Science* 5 (2014): 279–285; Erich Uhlmann, et al., "When Actions Speak Volumes: The Role of Inferences About Moral Character in Outrage Over Racial Bigotry," *European Journal of Social Psychology* 44 (2014): 23–29.

have strong emotional reactions to acts that we view as disgusting[139] or highly personal.[140] We automatically form ideas about character from our perception of extra-moral information, such as body language,[141] facial features or expressions,[142] and vocal inflection.[143] It is also widely understood that gossip plays a major role in shaping our understanding of who others are, especially where their character is concerned.[144] Given a limited amount of time to interact with others, we greatly benefit from the opinions of those we already trust, and even from information provided by strangers—as when we consult online reviews and testimonials.

This strong "person-centered" approach to moral psychology is relatively new, and it forces a reframing not only of the classic emphasis on moral rules, but also of psychologists' understanding of the role of so-called theory of mind (ToM) computations in moral judgment. Let us accept for the moment the original definition of ToM as "an ability to impute mental states to [oneself] and others,"[145] a definition I will nuance shortly. A person-centered approach

---

[139] Haidt, et al., "is it Wrong to Eat Your Dog?" 613-628.

[140] Greene, et al., "An fMRI Study of Emotional Engagement in Moral Judgment," *Science* 293 (2001): 2105-2108.

[141] David DeSteno, et al., "Detecting the Trustworthiness of Novel Partners in Economic Exchange," *Psychological Science* 23 (2012): 1549 –1556.

[142] Moshe Bar, et al., "Very First Impressions," *Emotion* 6.2 (2006): 269-278.

[143] Nalini Ambady, et al., "Surgeons' Tone of Voice: A Clue to Malpractice History," *Surgery* 132 (2002): 5-9.

[144] Robin Dunbar, "Gossip in Evolutionary Perspective," *Review of General Psychology* 8.2 (2004): 100-110.

[145] Heinz Wimmer & Josef Perner, "Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception," *Cognition* 13 (1983): 104. Adapted from David Premack & Guy Woodruff, "Does the Chimpanzee Have a 'Theory of Mind'?" *Behavioral and Brain Sciences* 4 (1978): 515-526.

to understanding our moral psychology helps explain, at a deeper level, why ToM inferences and moral rule monitoring matter for socio-moral cognition. Why do we constantly make inferences about others' mental states (such as beliefs, intentions, and emotions) and monitor their compliance with moral rules? An obvious answer is that, as highly social beings, we need to be able to predict their behavior and know, in turn, how we ought to behave in the most appropriate, kind, or otherwise successful way. However, consider that this suggests a further question: why should we expect to be able to predict the behavior of others based on such inferences concerning their knowledge, their emotional states, and the moral rules that are in effect? Much of this information would be useless if we did not already have a working hypothesis about what kind of person they are. For instance, we must already suspect that, given a certain belief, a person is likely to do one thing, whereas without this information, she will more likely do another thing—otherwise, inferences about her beliefs would be useless. Similarly, we must have some understanding of the kind of person she is before we can make use of our inferences regarding her emotional states—minimally, we must believe that she is capable of emotion at all—and only then can we usefully hypothesize about how her emotional state will affect her behavior. Thus, the meaningfulness of mental state inferences depends structurally upon a prior theory of person.[146] A slightly different relation obtains between person perception and our attention to rules. We can monitor rule-compliance without first forming hypotheses about the people whose behavior we are monitoring; thus, our attention to moral rules has a structural independence from person perception. Nevertheless, it is also true that, inasmuch as we want to be able to predict others' adherence to rules, it is useful to have a model of who they are; and

[146] Something along these lines is proposed by the philosopher Evan Westra, "Character and Theory of Mind: An Integrative Approach," *Philosophical Studies* (Advance Publication: 2017).

others' rule-breaking or compliance can help us update our working person-model. In this way, our attention to rules is similar to our attention to mental states: they both draw upon preexisting theory regarding what and who others are.[147]

As it relates to moral phenomena, this working hypothesis of what / who others are might be called a "theory of character" (ToC).[148] Such a "theory" may often be no more than a set of unconscious assumptions, which we update when others' behavior contradicts our expectations. Placing emotional and behavioral cues within the appropriate context, we form "snap judgments" about character and then update those judgments inductively and deductively. As such, our ToM and deontological reasoning abilities both draw upon a preexisting ToC and feed information into an ongoing inferential process of character attribution, wherein hypotheses about the character traits of others are reinforced, modified, or abandoned as new information is taken in. A growing body of evidence suggests that this process is a central focus of socio-moral cognition. For instance, as I indicated in chapter 2, studies show that moral "character" is the

[147] Perhaps this is why young children treat moral rules as being objective and unbreakable. Without a very sophisticated model of who others are, it may be expedient to assume that moral rules will be followed, rather than attempting to generate more nuanced hypotheses. As we mature, however, we come to understand that obeying or breaking rules is up to individuals whose choices reflect who they are. This could be true even if the so-called situationist critique is true, since we tend to encounter people in consistent contexts, which allows us to predict future behavior. I address "situationism" again in an endnote at the end of this chapter.

[148] This would be a "theory" based on the same criteria Premack and Woodruff initially adduced for calling mental state attributions a "theory of mind." Namely, like a mental state, character "is not directly observable," and it "can be used to make predictions" (Premack and Woodruff, "Does the Chimpanzee Have a Theory of Mind?" 515).

most important thing to which we are attuned when we observe other people,[149] and it is also the quality we perceive as being most essential to who we are.[150]

The person-centered approach helps explain the relevance of ToM computations, but it does not replace ToM. Nor can character attribution be treated as a subtype of ToM computation. Instead, character and mental state attribution must be carefully distinguished, precisely so that their complementarity can be properly appreciated. In practice, this distinction is harder to make than one might suppose at first glance, and imposing it will mean critiquing and redefining what has historically been considered the ToM domain. At a conceptual level, however, it is pretty simple. The distinction comes down to the difference between a mental "state" and a mental "trait." Mental states are relatively context-dependent. Our knowledge and beliefs change as a function of the information we have; our intentions change as we pursue different goals; and our emotions change with circumstances and moods. These are all "states" analogous to the different states (solid, liquid, or gas) that water takes at different temperatures. By contrast, character "traits" are more like the lasting chemical properties of H_2O, which stay constant across such physical transformations.

While this distinction is clean in principle, the reality is much messier. Consider this question: are "intentions" mental states? Although psychologists have claimed that they are, "intention" is a fuzzy concept, and the answer may depend on how the word is used. On one hand, "intention" might specify what someone is trying to accomplish here and now. In this

---

[149] Goodwin et al., "Moral Character Predominates," 148-168. I will discuss the "situationist" critique of character, along with the so-called fundamental attribution error in chapter 5.

[150] Strohminger & Nichols, "The Essential Moral Self," 159–171; Strohminger & Nichols, "Neurodegeneration and Identity," 1469-1479.

sense—the sense of a desire to accomplish an immediate goal—it seems clear that intention is a mental state. When we have accomplished the goal we can focus our mental energy somewhere else, and when we do, our mental "state" changes. On the other hand, "intention" might also mean something more lasting and context-independent. If I ask a young man about his "intentions" toward my daughter—to use a worn trope—I am not asking about a transient "state" of mind but about something more lasting, which I expect to survive changes in passing desires, thoughts, and feelings. Such "intentions" might be better compared to the persistent chemical properties of H_2O than to its "state" as ice, water, or steam. Even as the young man's mental "states" change, I want to know if his "intentions" will remain in some sense "pure." Are such "intentions" mental "states," or am I really asking about more lasting "traits," more essential properties of who my daughter's boyfriend is? The same problem confronts us if we contemplate another category of mental states—"beliefs." Classic ToM studies of our ability to properly impute beliefs to others typically test our ability to use local contextual cues to draw logical inferences about what another person thinks. In these cases, the "belief" is inferred almost entirely from context, in view of only the most generic assumptions about the person holding the belief. But a "belief" may also be something more trait-like, such as the core convictions one carries over years or even an entire lifetime—something that expresses more of the specificity of who someone is, rather than the local circumstances they are in. We hold onto our core moral beliefs as other mental states change—even as our understanding of the world changes—and we understand ourselves to be essentially defined by the core moral beliefs that we hold: indeed, we feel that if such beliefs were to change dramatically, we would no longer be the same person.[151]

[151] Strohminger & Nichols, "The Essential Moral Self," 159–171; Strohminger & Nichols, "Neurodegeneration and Identity," 1469-1479; Larissa Heiphetz, "The Role of Moral

Are the latter kind of "beliefs" properly characterized as mental "states"? Maybe not. Finally, I have defined character traits as being relatively context-independent, but the reality is not perfectly clean. Instead, there is quite a bit of evidence that character "traits" like honesty and helpfulness are expressed somewhat inconsistently, varying with situational factors.[152] Thus, these "traits" may be more state-like than we tend to assume.

The conceptual and empirical ambiguity in this area means we must attend to the phenomena of interest, rather than getting too hung up on the words used to characterize these phenomena. We must recognize that some "beliefs" are relatively context-dependent and more state-like, while others are relatively context-independent and more trait-like. Ditto for "intentions." Even character "traits" are not always perfectly trait-like, but may vary with context. The crucial innovation of the person-centered approach to moral psychology is the recognition that our attributions of state-like beliefs, intentions, and feelings are always made in view of prior trait-like attributions of the capacities and characteristics of others. The trait-like attributions are why state-like attributions matter in the first place, why they have any meaning and usefulness. But this does not mean that ToM computations are *reducible* to person perception. Just as knowing the chemical composition of H_2O does not tell us whether we are dealing with a solid, liquid, or gas, knowing someone's character traits does not give us specific details about their mental states. And there is a lot of research suggesting that the ability to accurately attribute context-dependent beliefs, intentions, and feelings is an important contributor to socio-moral cognition. ToM still matters.

Beliefs, Memories, and Preferences in Representations of Identity," *Cognitive Science* 41 (2016): 744-767.

[152] Kwame Anthony Appiah, *Experiments in Ethics,* Cambridge: Harvard University Press, 2008, chapter 2.

For example, people with high-functioning autism (HFA) are believed to have selective deficits in their ToM capacity.[153] And there is evidence that these populations are less likely than typical populations to be lenient in moral judgments of unintentionally harmful actions,[154] presumably because people with HFA have difficulty performing the ToM computation that distinguishes accidental from intentional wrongdoing.[155] Our ToM capacities affect much more than how we reason about moral infractions, however. Developmental evidence suggests that there are strong links between the development of a theory of mind and the development of some fundamental features of our moral psychology. For instance, the first signs of self-conscious emotions like shame and embarrassment occur around the time a typical child is able to recognize him or herself in a mirror (around 21-24 months). These emotions rely on an awareness of oneself as the object of others' perception and judgment—and it can be reasonably argued that this awareness is a kind mental-state attribution. Concurrent with this emergence of self-consciousness, children begin to exhibit a sense of ownership, a concern with their reputation, and an increase in spontaneous helping and comforting behaviors between their 2nd

[153] Charlotte Montgomery, et al., "Do Adults with High Functioning Autism or Asperger Syndrome Differ in Empathy and Emotion Recognition*?" Journal of Autism and Developmental Disorders* 46 (2016):1931-1940; Isabel Dziobek, et al., "Dissociation of Cognitive and Emotional Empathy in Adults with Asperger Syndrome Using the Multifaceted Empathy Test (MET)," *Journal of Autism and Developmental Disorders* 38 (2008): 464–73.

[154] Moran, Joseph, et al., "Impaired Theory of Mind for Moral Judgment in High-Functioning Autism," *PNAS* 108.7 (2011): 2688–2692.

[155] Note that "intentionality" in these studies is state-like. Although one's inference about intentions in these morally ambiguous scenarios does feed into a process of character attribution, it only does so by way of initial prior state-like attribution of local intentions, made in view of only the most generic trait-like assumptions.

and 3rd year.[156] Shortly thereafter, a burgeoning ability to distinguish between beliefs and reality enables children to begin dissembling and lying. This latter fact is no mere developmental correlation but a logical necessity—before you can tell a lie, you must first appreciate that the person you are lying to is capable of believing something that is not really the case—that is, you must be able to distinguish beliefs from reality. Buttressing this deductive argument, there is striking empirical evidence of a link between the development of ToM and the emergence of lying. Children who are better at classic ToM tasks lie more frequently and are better able to maintain their lies in the face of questioning than children with less advanced ToM abilities.[157] And a recent study even showed a direct causal relation between the development of ToM abilities and lying behavior, as children who had not demonstrated an ability to lie received training on ToM tasks and subsequently became adept and prolific liars (compared to a control group that received training in non-ToM reasoning).[158]

Such evidence suggests that ToM computations are of fundamental importance to humans' moral psychology. Our ability to lie, to feel shame, even to engage in appropriate helping behaviors depends upon the development of ToM capacities. Moreover, the current

---

[156] Philippe Rochat, "Layers of Awareness in Development," *Developmental Review* 38 (2015): 141; Philippe Rochat, et al., "Fairness in Distributive Justice by 3- and 5-year-olds Across Seven Cultures," *Journal of Cross Cultural Psychology* 40.3 (2009): 418; Zahn-Waxler, Carolyn, Marian Radke-Yarrow, & Elizabeth Wagner, "Development of Concern for Others." *Developmental Psychology* 28.1 (1992): 126-136.

[157] Victoria Talwar and Kang Lee, "Social and Cognitive Correlates of Children's Lying Behavior," *Child Development* 79.4 (2008): 866-881. However, this study did not control for factors such as general intelligence and other contextual factors that might contribute to lie-telling.

[158] Xiao Pan Ding, et al., "Theory of Mind Training Causes Honest Young Children to Lie," *Psychological Science* 26.11 (2015): 1812–1821. Note that lying was the winning strategy in the game these children were playing, and so we shouldn't presume that their increase in lying was a morally negative thing.

evidence—especially studies of people on the autism spectrum, whose ToM capacities seem to be selectively impaired even when their general intelligence is high—supports the idea that ToM is a special kind of computation, which cannot be reduced to some domain-general rational competency. Nevertheless, it is also true that ToM computations are meaningful and useful primarily in view of prior attributions of personal traits and, in the moral domain, especially considerations of moral character.

All this is interesting and relevant, but it continues to fall short of answering the question that we have been missing from the beginning. We have answered the question why humans are so obsessed with personal traits and mental states. The person-centered model claims that all these inferences matter in view of our need to know who we can rely on in various cooperative and competitive situations and how we can behave in ways that are maximally appropriate, effective, kind, and so on—always in view of our working models of (generically) what and (specifically) who others are. But accepting these practical answers—that these forms of knowledge help us successfully navigate our social world—we are left with a final question: how do we distinguish such practical success from moral success? What is the aim of moral cognition, as opposed to any endeavor for which our inferences about others' traits and states might be useful? What is morality all about? I'll put off answering this question for just a few pages more.

Moral Feelings

Moral rules, personal traits, and mental states are the basic objects of socio-moral cognition, according to my interpretation of the current literature, although I have been hinting that there is some missing term in this equation of moral cognition, which I will soon fill in.

First, however, my critical review of current research would be woefully incomplete if I failed to discuss the role of feelings and intuitions in moral judgment. I will start with the most famous of such feelings. Work from Daniel Batson,[159] Martin Hoffman,[160] and Frans de Waal[161] over the past couple decades has revived the old "sentimentalist"[162] emphasis on our ability to feel empathy, showing not only how important this capacity is for successfully navigating the socio-moral landscape, but also how evolutionarily deep empathy is—for instance, de Waal argues that many mammals and birds exhibit empathy for their young and suggests that this parent-child relationship is the evolutionary seed from which all empathic feelings and behaviors ultimately sprouted.[163] People who exhibit low levels of affective empathy, such as people diagnosed with trait alexithymia (characterized by emotional deficits) or psychopathy, or patients with vmPFC damage, tend to rely more on rationalistic, "utilitarian" forms of moral judgment than people with typical levels of empathy.[164] But this rationalistic moral style is not necessarily a good thing. In the case of psychopathy, for instance, there is a striking and well-established correlation with criminal behavior—especially violent crime—and it is widely acknowledged that this

---

[159] Daniel Batson, *The Altruism Question: Toward a Social-Psychological Answer* (NJ: Lawrence Erlbaum, 1991).

[160] Martin Hoffman, *Empathy and Moral Development* (Cambridge: Cambridge University Press, 2000).

[161] Frans de Waal, *The Age of Empathy: Nature's Lessons for a Kinder Society* (New York: Harmony Books, 2009).

[162] So-called sentimentalist philosophers Adam Smith and David Hume emphasized the role of feelings in motivating moral judgment, back in the 18th century.

[163] de Waal, 2008, 282.

[164] Patil, Indrajeet, and Giorgia Silani. "Reduced Empathic Concern Leads to Utilitarian Moral Judgments in Trait Alexithymia." *Frontiers in Psychology* 5.501 (2014): 1-12.

correlation owes largely to psychopaths' deficiency in affective empathy, even as their ToM and other cognitive abilities are often spared.[165]

While I have been emphasizing the feeling-component of empathy, empathy actually involves two distinct components—both a mental-state attribution and a responsive feeling. In order to exhibit empathy, I must first perceive or understand at some level what someone else is experiencing; only after such a ToM computation, for me to exhibit empathy in the classic sense, I must have an affective or emotional response and thereby participate in the experience of the other.[166] Empathic feelings are far from the only important moral sentiments, however, and many of the other processes involved in socio-moral cognition also have strong correlative feelings. One's inference that they have been treated unjustly is likely to be accompanied by a burst of indignation. A feeling of disgust may accompany a negative judgment of others' conduct or character. We feel warmth or elevation while witnessing kind or noble deeds. The existence of such moral feelings has been recognized for a long time, however. What has been less clear is what role such feelings play in moral cognition and action.

---

[165] Kent Kiehl & Morris Hoffman, "The Criminal Psychopath: History, Neuroscience, Treatment, and Economics," *Jurimetrics* 51 (2011): 355–397; Harma Meffert, et al., "Reduced Spontaneous but Relatively Normal Deliberate Vicarious Representations in Psychopathy," *Brain* 136.8 (2013): 2550-2562.

[166] Readers might want to raise a philosophical objection about the possibility of having accurate understanding of another's experience. Is my empathy for you created by my ability to truly understand what you are going through—your mental state, in this sense—or am I simply imagining how I would feel if I were in your situation? If I am only doing the latter, I may draw inaccurate inferences about your mental states (indeed, mental state attribution may never be perfectly accurate). This objection is noted. But the claim is not that mental state attributions must be perfectly accurate in order to be considered ToM inferences. Insofar as I attribute an experience to you—even when I am only imagining myself in your shoes—I am performing a theory of mind computation—an imperfect one, yes, but a ToM computation all the same.

Are such "moral sentiments," as they have been called,[167] mere epiphenomena, accompanying certain cognitive processes but doing nothing to determine moral reasoning and action itself—or worse, actually hindering moral deliberation, as many philosophers have suggested? Or do moral feelings play a more central and salutary role in moral cognition and action? Neurobiologist Antonio Damasio famously analyzed a number of patients with damage to their ventromedial prefrontal cortex (vmPFC), an important hub for emotion regulation—comparing these patients to the mid-nineteenth century railroad construction foreman Phineas Gage, who famously survived after a work accident sent a three-and-a-half-foot iron rod through his skull. Like Gage, many of these patients retain the majority of their rational functions and even ace many psychological tests, including Lawrence Kohlberg's famous moral dilemmas; yet they typically engage in self-destructive and immoral behavior in their actual lives, finding it difficult to follow through on commitments or behave in socially appropriate ways without their old emotional cues.[168] Thus, Damasio argued that, far from being a hindrance to moral rationality, socio-moral emotions are a central part of functional moral cognition.

In view of the above evidence, I accept that feelings play a central role in socio-moral cognition and are not mere epiphenomena. We see distinctive changes to our moral psychology when affective / emotional faculties are impaired. While people with ToM impairments have trouble performing rational moral computations, like moral discounting of accidental harms, people with emotional impairments tend to rely on rationalistic forms of moral judgment and

---

[167] See Adam Smith, *The Theory of Moral Sentiments* (Oxford: Oxford University Press, 1976 [Originally published in 1759]).

[168] Antonio Damasio, *Descartes' Error: Emotions, Reason, and the Human Brain* (New York: Avon Books, 1994).

may have a variety of social problems, including a higher tendency to criminal behaviors and higher rates of recidivism.

Moral Pluralism

Any model or theory of socio-moral cognition must be consistent with the literature discussed above, recognizing that the human mind is highly attuned to moral rules, character, and mental states, and that socio-moral cognition is facilitated by affective and emotional cues. Many theories might fit within these parameters, however. For instance, there are there is an ongoing debate between so-called moral pluralists and monists over the question of how complex our theories need to be: are there multiple irreducibly distinct moral values, or can all moral values ultimately be explained in terms of a single moral imperative? Both monistic and pluralistic theories should be consistent with the above literature, but no existing monistic or pluralistic theory can explain the polyphonic dynamics of Raskolnikov's confession.

Early models of moral psychology tended to be monistic, trying to explain all moral computations in terms of a single value. For instance, both Piaget and Kohlberg saw their theories as being essentially about what Kohlberg called "justice reasoning."[169] However, Kohlberg's former student, Carol Gilligan pointed out that some humans evidently have moral concerns that are not reducible to justice. Some contemporary monists, such as Kurt Gray, argue that morality is all about interpersonal harm.[170] Haidt has led the way in critiquing every form of

---

[169] Kohlberg, *The Psychology of Moral Development,* 215-217.

[170] Chelsea Schein & Kurt Gray, "The Unifying Moral Dyad: Liberals and Conservatives Share the Same Harm-Based Moral Template," *Personality and Social Psychology Bulletin* 41 (2015): 1147–1163; Kurt Gray, Chelsea Schein, & Adrian Ward, "The Myth of Harmless Wrongs in Moral Cognition: Automatic Dyadic Completion From Sin to Suffering," *Journal of Experimental Psychology: General* Advance Online Publication (2014).

moral monism, including Gray's. Drawing on anthropological literature indexing and categorizing socio-moral systems across diverse societies, Haidt and colleagues identify five or six irreducible value-domains, "moral foundations," which they believe organize human moral intuitions in advance of experience. Moral intuitions related to these basic values turn up in diverse cultures around the world, and Haidt hypothesizes that these intuitions are facilitated by cognitive adaptations to recurring challenges faced by our evolutionary ancestors. For instance, in order to provide for their children, our (very distant) ancestors evolved an ability to feel empathy for their offspring. Over time the ability / tendency to feel warmth, compassion, and love was eventually extended to others outside the parent-child dyad. This "care / harm" foundation now predisposes us to develop certain moral "intuitions" about the treatment of babies and non-human animals, victims and offenders to whom we aren't related, and so on. Other "foundations," including fairness / cheating, loyalty / betrayal, authority / subversion, and sanctity / degradation, are also proposed to organize our moral development in advance, though our unique cultural environment determines which foundations we "build" upon—and to what extent we do so—and which we repress.

I basically buy Haidt's argument, but I'm not going to get into the many empirical studies through which this battle over monism vs. pluralism continues even today. My reason is that the debate itself rests upon problematic assumptions about the nature of morality and moral cognition. Even though I am convinced by Haidt's argument that there are more than one irreducibly distinct moral values, the attempt to define the moral domain in terms of moral values is circular, and we must ask what it is that makes a moral value moral. This originary source of morality, which neither Haidt nor his opponents nor anyone in the field of moral psychology has

uncovered, is also the source of the polyphonic dynamics of my Dostoevskian model of moral cognition. While distinct moral values can offer a modest contribution to moral polyphony, a theory of value pluralism cannot account for the real-world complexity of Raskolnikov's confession. By the end of this chapter, you will know why.

Implicit Cognition

Since chapter 1, I have allied myself with Haidt's Social Intuitionist Model of moral psychology, which proposes that "intuitions" are much more decisive for moral judgment than rational deliberation, and that the rational deliberation we *do* engage in is typically done for the purpose of justifying our actions to other people—after we have already made up our minds. However, I must qualify my alliance with Haidt's intuitionism. Moral "intuitions," for Haidt and colleagues, are "the sudden appearance in consciousness, or at the fringe of consciousness, of an evaluative feeling (like–dislike, good–bad) about the character or actions of a person, without any conscious awareness of having gone through steps of search, weighing evidence, or inferring a conclusion."[171] Haidt gives evidence suggesting that rational justification in support of such moral intuitions tends to be post hoc, with people typically making rapid, intuitive decisions and then searching for compelling justifications for these decisions only after their initial judgment is challenged by others.[172] Haidt thus argues for the importance of a particular kind of implicit (i.e., unconscious) cognitive process—that which gives rise to a moral "intuition"—while criticizing

---

[171] Graham, et al., "Moral Foundations Theory," 66. Adapted from Haidt, "The Emotional Dog and its Rational Tail," 814.

[172] Haidt, "The Emotional Dog," 814-834; Jonathan Haidt & Fredrik Bjorklund, "Social Intuitionists Answer Six Questions About Morality," In *Moral Psychology*, *Vol. 2*, ed. Walter Sinnott-Armstrong (Cambridge: MIT Press, 2008), 181-217; Haidt, et al., "is it Wrong to Eat Your Dog?" 613-628.

the emphasis of many moral psychologists and philosophers on explicit and deliberative forms of moral reasoning. The problem is, this opposition between moral intuitions and deliberative moral reasoning leaves out a whole bunch of implicit cognitive processes that are neither deliberative nor "intuitive" in Haidt's sense. Let us consider some of these ignored processes.

Asking participants about the possibility of performing immoral, irrational, improbable, and impossible actions, psychologists Jonathan Phillips and Fiery Cushman showed that when time for deliberation was restricted to 1.5 seconds or less, participants tended to think that immoral actions were impossible; whereas, when they had more time to reflect, they were more likely to recognize the possibility of performing immoral acts. Crucially, this effect was observed specifically when participants were reasoning about scenarios with moral content, as opposed to scenarios that did not have moral content.[173] This finding suggests that humans' understanding of right and wrong implicitly constrains our expectations for how others *are able* to behave, thus limiting the amount of information we have to sift through to predict others' behavior. This is just the sort of non-rational implicit process that someone like Haidt might want to invoke in arguing against rationalism. However, since this implicit cognitive process has no associated feeling, it is not a Haidtian "intuition."

---

[173] Jonathan Phillips & Fiery Cushman, "Morality Constrains the Default Representation of What is Possible," *PNAS* 114.18 (2017): 4649-4654. This work extends Piaget's observation that young children often seem to view moral rules as objective constraints on action. It also indicates that similar assumptions continue to structure adults' implicit representation of reality. However, it seems to contradict Piaget's model of stages, according to which each new stage of development fundamentally reorganizes our way of seeing things, so that we lose the representational content of earlier stages. Instead, the continuity between our earlier-stage representations and implicit representations in adulthood is more consistent with the "onion" metaphor used by Philippe Rochat to describe the "layers of awareness" that we develop over time, with each new layer adding something new to our representations of reality but not destroying or fundamentally reorganizing previous layers (see "Layers of Awareness in Development," 122-145).

This point about the importance of implicit cognition applies to moral cognition broadly, not just within the domains of Haidt's moral foundations. Just as many of our judgments about moral rules involve implicit cognition, so too do attributions of character and mental states. For instance, we engage in implicit ToM processing when we "intuitively" respond to the perceived emotional state of another person, or when we automatically respond to the perceived needs of others, which even very young children do.[174] Indeed, as with most cognitive processing, I consider it likely that the majority of our attributions of others' intentions, feelings, and even certain kinds of beliefs[175] happen implicitly, without the need for conscious reflection. ToM perceptions are sometimes accompanied by feelings, as in the case of affective empathy. But much of the time we simply make attributions without feeling much of anything. Thus, many of the implicit ToM processes that shape moral cognition do not qualify as Haidtian intuitions. A similar observation holds for character attribution. While our understanding of others' character may ultimately involve reasoned consideration of facts, I already described how character attribution begins as a rapid implicit process. Such initial representations of character may or may not be accompanied by feelings. Implicit character attributions that do not give rise to feelings would not qualify as Haidtian intuitions, but that does not mean that they are irrelevant to moral cognition. Finally, consider that feelings can also be associated with explicit cognitive processes. For instance, if I dwell on the fact that I have been wronged in some way, my conscious deliberation is likely to be animated by a strong feeling of indignation.

---

[174] Carolyn Zahn-Waxler, et al., "Development of Concern for Others." *Developmental Psychology* 28.1 (1992): 126-136.

[175] For instance, we might engage in many forms of misdirection, hiding, pretending, and other forms of deception automatically—all of which require some basic attribution of what the other person perceives and thus, in this local context-dependent sense, what they "believe" (see my discussion of context-dependent vs. context-independent "beliefs" in the section on ToM).

The above evidence suggests a theoretical revision for Haidtian intuitionists. If we focus on "intuitions" that involve both implicit cognition *and* feelings, we ignore much cognitive content that is interesting and important to socio-moral cognition: namely, the many forms of implicit cognition that don't give rise to noticeable feeling-content but which nonetheless affect moral perception, judgment, and action, and the evaluative feelings that arise in the context of explicit moral reasoning. Thus, we should replace the imperfect opposition between deliberative moral reasoning and Haidtian moral "intuitions" with the more expansive opposition between explicit and implicit moral-cognitive processes. And we should appreciate that both implicit and explicit moral-cognitive processes can be domain-specific or domain-general, and that both processes can involve strong feelings or no feelings at all.

Conclusion on Socio-Moral Cognition

It is well established that rational inferences about moral rules and others' mental states are key for socio-moral cognition, and evidence of the past couple of decades also indicates that feelings play a central role—since, when our ability to feel normal social emotions is impaired, so is our ability to successfully navigate our socio-moral environment. More recently, it has become evident that character is also a central focus of social cognition, and I have argued that a fundamental concern with character provides an important reference point for computations about moral rules and mental states—since, in the absence of a working model of others' character, inferences about mental states and moral rules would be less predictive of their behavior. All of these forms of cognition may happen either implicitly or explicitly, but I am inclined towards the view that the bulk of our social and moral cognition is implicit, happening

below the level of our conscious awareness. And I add the caveat that implicitness is a very broad category, not restricted to Haidtian "intuitions."

Taken together, all this complexity does not constitute moral polyphony. These are the tools we use to navigate our socio-moral worlds, but the work of morality that we use all these tools for might be rather straightforward, as moral monists propose. In such a simple moral world, we might develop moral competency in the same way that we develop competency in many other domains, becoming increasingly skilled at moral cognition and behavior, increasingly moral, with practice. However, moral cognition is much more complex than philosophers and psychologists have imagined. Value pluralism—such as that proposed in Haidt's Moral Foundations Theory—is one way in which a polyphonic psychological dynamic might arise. But such pluralism doesn't account for the pervasiveness of polyphony in our everyday lives. We are not forced to negotiate between values of justice and compassion, for instance, on a constant basis. But my claim is that polyphony is the rule in everyday moral cognition, even though we are largely unaware of it. And, indeed, it would be the rule even if moral cognition were utterly monistic. I devote the rest of this chapter to making this case.

In what follows, I claim that moral polyphony primarily arises from something more primordial than values, which the debate over moral monism vs. pluralism obscures. This primordial source of moral polyphony is also the fountainhead of morality itself. Tracing the polyphonic dynamics of moral cognition in the real world thus requires us to identify the transcendental conditions of morality and moral cognition—a task that, as far as I am aware, has never been performed.[176]

[176] We will see that specifying the transcendental conditions of philosophical ethics (so-called metaethical principles) is a secondary task that does not give us the transcendental conditions of morality in the purely descriptive sense necessary for a science of moral psychology. Thus, for

Existential Framing

Introduction: Parochialism, Impartiality, and Sorge

A few pages back, I claimed that although people are *able* to adopt a (relatively) impartial perspective when asked to resolve moral dilemmas, this objectivity is not our normal position when making moral judgments or acting in moral or immoral ways. In reality, our personal investments in projects and our relationships with people, places, creatures, objects, and ideas shape how we see right and wrong. Moral cognition tends to be parochial, I claimed. Parochialism makes a lot of sense. Indeed, the fact that we ever concern ourselves with what is *really* best, as opposed to what is best for ourselves and those close to us, has been seen as a paradox by evolutionary theorists, and is at the very least a perennially interesting puzzle. Nevertheless, it must be acknowledged that many of us *are* invested in universal ideals, such as justice and truth. As such we find ourselves in a paradoxical moral position, as our commitment to universal ideals vies with our innate parochialism. What if I told you that a single property of our minds lies at the origin of both our desire to be impartial and our parochial moral tendencies? Not that these inclinations are essentially the same—they are diametrically opposed—but that they each begin from a common potentiating source. This property—which has yet to be properly appreciated in its significance for moral psychology—affects most of the other forms of cognition we have so far discussed and potentiates the polyphonic dynamics of my model.

One name for this property is "Sorge," a crucial concept in *Being and Time*, the introductory volume to philosopher Martin Heidegger's unfinished exploration of the meaning of

---

instance, Kant's *Groundwork for the Metaphysics of Morals* does not give us the transcendental conditions for morality but only performs the secondary task of exposing the transcendental conditions for what Kant considers a *normative* moral system.

being. Sorge means "concern," but in *Being and Time* it does not simply express a kind of worriedness or anxiety, as might be supposed. In this work, it is typically translated as "care," but Heidegger's concept of "Sorge" does not mean compassion or empathy either. Indeed, in *Being and Time*, "Sorge" does not denote any particular feeling or mood at all. Rather, it refers to our distinctive mode of encountering and interpreting things in the world *as meaningful*.[177] When we love or hate, when we feel attraction or disgust, when we are frustrated or satisfied, even when we don't feel anything in particular but simply navigate our world in the usual way, we always respond from within a world that matters to us, and Heidegger saw that this meaningfulness of things is more fundamental to our ontological understanding than the objective properties of things.

Take the keyboard keys on the keyboard on which I am currently typing. Although I could define the keys objectively, in terms of their measureable physical properties—and I could set objective parameters for inclusion of objects within the category of keyboard keys—before I do any of this the keys already have an instrumental meaning, which is that they are for typing. And typing is meaningful in the context of writing or coding, and these activities are important for e-mailing, creating experiments, composing a dissertation, and any number of other valued endeavors. And crucially, my knowledge of the objective properties of my keyboard is only relevant because of this prior interpretation of the keys and keyboard as being for typing and this prior understanding of why typing matters.[178] For instance, the size of the keys matters because

---

[177] This is not restricted to semantic meaningfulness. I am using the term in a broad and colloquial sense.

[178] Heidegger calls this the "as" structure of interpretation. See Martin Heidegger, *Being and Time*, trans. Joan Stambaugh & Dennis Schmidt (Albany: SUNY Press, 2010 [original 1927]) 144, *Sein und Zeit* 148-149. I am indebted to Taylor Carman's exposition of the "as" structure in

keys must be of the proper size to facilitate typing. A key that is too big or too small for this purpose is a bad key. But until I understand the key as being for typing—an activity that itself only matters because of a greater context of meaningful activity—I cannot have any idea which objective properties matter. And this is because the fundamental ontological meaning of a key is not reducible to any objective property or properties, however specific and detailed such objective description may be.

Why do we do it? Why does it matter to us? What is its significance? These are the kinds of questions that are asked only by beings whose understanding is characterized by Sorge—and these questions are always asked in view of ends that are intrinsically valuable and, as such, worthy of pursuing. We ask instrumental questions too—How do we accomplish it? How can it be quantified? How can we be confident it is true? How does it work? What are its physical properties? How can it be classified? But these instrumental questions are always asked in view of the prior existential questions above, and are only meaningful in this connection. Thus, one of Heidegger's great insights was that our ontological understanding of our world, and of our place within the world, is grounded first and foremost in Sorge—not in our ability to draw rational inferences about the objective properties of things.

Why does this matter for moral psychology? A consequence of the grounding of ontological understanding in Sorge is that existential judgments of *what things are* tend to co-emerge with normative understanding of *how they ought to be*. The keyboard key that is too big or too small, too spiky or sticky or brittle, is a bad key—not because there is anything inherently bad about relative size, stickiness, spikiness, or brittleness, but because the key is for typing, and its features ought to conduce to typing (bearing in mind that typing is meaningful in view of

---

Heidegger's Analytic: Interpretation, Discourse, and Authenticity in Being and Time (Cambridge: Cambridge University Press, 2003) 18-22.

expressing, sharing, criticizing, collaborating, and so on). A similar existential dynamic obtains in realms of social and moral normativity. For instance, if I call you my friend, this ontological statement is normatively conditioned. Insofar as my actions do not conduce to friendship, in the myriad ways in which I may so fail, I *am* a bad friend—if my actions are offensive enough, I may cease to be a friend altogether. I may likewise fail *as* a brother, a citizen, a soldier, a son, a teacher, a lover, even as a human being. I may fail other people, animals, plants, sacred places and objects, the earth, my ancestors, or even my future self. I may fail in my parochial involvements with people, creatures, places, and ideas that matter to me: *my* friends, my pets, my home, my religion. And I may also fail with respect to my ideals of impartiality, justice, and universal rights—ideals that are *also* meaningful to me. Thus, I may strive to *be* impartial and fair, just as—in being a good friend—I am partial to my friends. In either case, it is my meaningful involvement with things that gives rise to my sense of how I ought to be. Thus, as I indicated above, both parochialism and impartiality are grounded in a common property of my nature—Sorge.

Because our understanding is structured by Sorge, we don't simply observe the objective features of things but typically interpret each person, place, object, or idea as good or bad, appropriate or inappropriate, satisfactory or unsatisfactory in some degree—and similarly, the ways in which we approach or *relate* to people, places, creatures, objects, and ideas are typically understood as either proper or improper, right or wrong. This pervasive normativity of our ontological understanding of things (and relationships) conditions all moral perception, judgment, and action. I call this fundamental normativity "existential framing," in honor of Heidegger's insight that Sorge structures (or frames) our understanding of existence. The meaning of existential framing and its relevance for moral psychology is the most crucial feature

of my polyphonic model. As will become clear, even if all the cognitive complexity described in the first twenty-five pages of this chapter were reduced to some simple moral-cognitive function, the dynamics of existential framing would still give rise to a moral polyphony.

Defining the Moral Domain in Terms of Existential Framing

If Sorge is a property of our nature that causes us to understand things as meaningful, existential framing is an outcome of this property—the reality that our interpretation of what things are is often normatively inflected. Within this normative dimension of ontological interpretation, there are distinct modalities in which we approach things, and each mode of approach has distinctive normative standards. For instance, we may approach something in an instrumental mode, with a normative standard of usefulness / uselessness. We may perceive or judge something in an aesthetic mode, with a normative standard of beauty / ugliness. We may consider something in an epistemic mode, with a normative standard of truth / falsity. And we may approach something in a moral mode, with a normative standard of rightness / wrongness or virtue / depravity. Any or all of these modes may apply to our ontological interpretations at any given moment.[179] They are all forms of existential framing. There is something special about the moral mode of existential framing, however. The moral mode is a meta-mode, a second-order normativity. It tells us how we ought to relate to that which has already been deemed good or bad. That is, moral norms are norms about how we ought to interpret things or behave in view of *prior* normative interpretations of beings. Without this prior normative framing, morality could

[179] This is not a comprehensive list. There may be other modes, governed by appetite, interest, and so on. Moreover, these modes are not cleanly distinct. From the inception of philosophy, philosophers have seen a higher unity in truth, beauty, and goodness, for instance. Though my polyphonic model contests any unfounded reduction of categories, I also recognize that there may be an aesthetic dimension to truth and goodness, an epistemic dimension to beauty, and so on.

never arise. It wouldn't make sense to say we ought to treat things with respect or love if nothing in existence were considered lovable or respectable. It wouldn't make sense to be obliged to defend or preserve things unless those things were already considered worthy of preservation. Thus, all moral norms rely on prior imputations of value to things, and before we can understand the relevance of existential framing for moral perception, action, and judgment, we must first understand non-moral modes of existential framing. This is what I will do in this section, beginning not with the moral mode of existential framing but with the instrumental mode, before moving on to phenomenologically define the domain of morality.

I begin by asking you to imagine a chair. Because our ontological understanding is grounded in Sorge, the chair is not merely a material object but is something with a normative quality. It is a good or bad chair, a comfortable or uncomfortable chair, an attractive or ugly chair, or whatever. How do we arrive at such normative interpretations of the chair? One way is to sit in the chair and judge it based on how comfortable or orthopedic it is. This is to judge the chair in view of how it fulfills its purpose as something designed for sitting on or in—a teleological and instrumental mode of interpretation. Humans are very much teleological beings. We often act with explicit goals in mind, and we interpret things in terms of how useful they are for our purposes. So, our normative assessment of a chair can arise from within a teleological frame. A chair that I cannot sit on or in, it might be said, is a *bad* chair. This is "bad" within the teleological interpretive mode of an instrumental existential frame.

What about when I walk into a fancy museum to find a beautiful chair suspended on cables from the ceiling, upside down? Or when I pass this installation to view a doll house with a perfect little chair inside? Am I to judge these as "bad" chairs simply because I cannot sit in them? Only if I have no idea what art is. Supposing I approve of these chairs, it is only possible

because my aesthetic framing of a chair as artistic display is distinct from an instrumental framing of a chair as something for my sitting on or in. Even if the museum chairs are very bad chairs in the latter sense, they may be quite good *as* expressions, representations, works of art. Thus, one's normative interpretation (or framing) of even a single object is not utterly simple. I may impose multiple ontological interpretations upon what is—objectively speaking—a single thing; and in this sense, a single object may have multiple essences, each judged within its own distinct normative frame. Here we have a chair that is bad for sitting in but good for displaying in a museum, instrumentally bad but aesthetically good.

Aesthetic judgment in this case is non-instrumental, non-functional, though beauty has long been recognized as occupying an ambiguous space, straddling the functional / non-functional divide. In chapter 2, I explored some of the peahen's anti-functional aesthetic preferences and drew an analogy to certain kinds of moral beauty that share the quality of the peacock's feathers, having negative-utility. And yet, I also acknowledged that functional features may be beautiful precisely because of their functional excellence. Kant described aesthetic judgment as an appreciation of purposiveness of form in the absence of a definite purpose for that form,[180] a formulation that acknowledges the aesthetic potential of functional form, but which holds that the aesthetic *mode of encounter* is non-instrumental or "disinterested."[181] Without getting into a nuanced debate over what aesthetic judgment *really* is,[182] we can simply recognize that our hypothetical museum chairs show that we can approach functional things in a non-instrumental way, and that our normative assessment changes along with our mode of

---

[180] Immanuel Kant, *Critique of the Power of Judgment,* 106, AA 5:221

[181] Kant, *Critique of the Power of Judgment*, 90-91, AA 5:204-5.

[182] Personally, I think there are radically different kinds of aesthetic judgment.

approach. A chair that is functionally useless and thus instrumentally bad may nevertheless be beautiful and thus aesthetically good. Again, we must recognize that, in a fundamental ontological sense, a single object may take on multiple essences. What it *is,* and how it is normatively judged, depends on how we approach it, what meaning we ascribe to it.

Another non-instrumental manner of interpretation is given by the moral approach to things. In treating moral existential framing, I will round out the definition of morality that I began in chapter 1 and fulfill my promise to phenomenologically define morality and moral cognition. Back then, I acknowledged that morality is a special kind of normativity: moral rules are seen as more universal and less context-dependent than merely conventional rules, and moral character is seen as core to our identity. Moral cognition is understood by some to be exclusively oriented towards justice, harm, and universal rights, while others would like to expand this domain to include things like loyalty, authority, and sanctity. As I said before, this debate over how to best operationalize the moral domain suffers from a conspicuous theoretical absence: we have no principle for distinguishing moral values from non-moral values. And no empirical evidence or ultimate evolutionary rationale can stand in for the principle that is missing. In chapter 1, I promised that I would eventually round out these definitions with a phenomenological account of how the moral domain arises through a process of existential framing. So here goes.

I begin my phenomenological account by interpreting the "respect for persons" formulation of the "categorical imperative," Kant's famous Uber-rule for any possible moral system, which he proposed as a guiding principle for distinguishing right from wrong in each particular case: "Act so that you use humanity, as much in your own person as in the person of

every other, always at the same time as end and never merely as means." [183] Kant arrives at this version of the categorical imperative by distinguishing free rational beings (such as humans) from the rest of nature. In contrast to objects, which may be framed instrumentally, a person is, in Kant's famous phrase, an "end in itself."[184] People have a rational faculty and can use it to conceive and pursue their own idea of the good,[185] says Kant, and so it would be totally irrational (normatively bad, within an epistemic frame) to treat them as if they were a mere instrument for our own pursuits.[186] Recognizing others as ends in themselves thus restricts us—insofar as we are rational—from treating others as mere instruments. And this restriction has far reaching moral consequences.[187]

In the interests of showing how the domain of morality emerges through existential framing, I would first like to draw attention to the fact that Kant's entire argument rests upon a radical procedure of existential framing. The "respect for persons" version of the categorical

---

[183] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, ed. & trans. Allen Wood (New Haven: Yale University Press, 2002 [originally published in 1785]), p. 46-47, AA 4:429.

[184] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, ed. & trans. Allen Wood (New Haven: Yale University Press, 2002 [originally published in 1785]), p. 45, AA 4:429. Colloquially, we say you shouldn't "objectify" people. Objects are understood instrumentally, in terms of how may they subserve our own goals, but people should be understood differently.

[185] Fortunately, thinks Kant, all rational beings must arrive at the same idea of the good. *Groundwork,* 70-71, AA 4:454.

[186] Similarly, for Aristotle, who proposed a teleological interpretation of virtue, a good person is not defined solely by what they contribute *to us* or to our community but by their ability to achieve Eudaimonia, their own happiness, *their own* highest end. See Aristotle, *The Nicomachean Ethics,* ed. E. Capps, T.E. Page, & W.H.D. Rouse and trans. H. Rackham (London: William Heineman, 1934) .

[187] Kant, *Groundwork*, 46-47, AA 4:429. Humanity means "rational being" and is not in principle restricted to the human species. Notice that Kant's grounds his moral argument in existential framing: an ontological claim of what is "rational" carries implications of normative goodness, within an epistemic frame, which subsequently gives rise to a moral responsibility.

imperative is simply a statement of what Kant views as the moral implications of three ontological claims: first, that freedom of will is requisite for something to be intrinsically good, or "good in itself"; second, that such freedom is only possible for rational beings; and third, that humanity is essentially defined by such rationality and freedom.[188] Insofar as we are rational, according to Kant, we will recognize these truths and treat others as ends in themselves—and every moral rule can be derived with reference to the imperative implicit in this recognition. Kant appreciates that before we can have moral norms, we must first have some non-moral idea of intrinsic goodness / badness for moral norms to be about, and his framing of humanity as a rational "end in itself" supports a moral principle guiding how we ought to treat people, compared to non-human things. In Kant's view, whereas it is appropriate to perceive, judge, and use "things" purely in view of our own ends, reason tells us to respect the rational capacity of others and to treat them as ends in themselves.[189]

The recognition that morality is a second-order normativity concerned with what has already been normatively judged as intrinsically good (or bad) is extremely important. But Kant does not thereby arrive at an adequate definition of the domain of morality, which we could apply to a model or theory of moral psychology—for one simple reason. Kant is trying to distinguish rationally valid ideas of intrinsic goodness from ideas he considers irrational and invalid. His method is thus totally opposed to the empirical method of psychology, and it is guaranteed to yield a narrower domain of "morality" than that which would be adequate for the

---

[188] Kant, *Groundwork,* 45-47, AA 4:428-429.

[189] This doesn't mean that we don't evaluate other people in instrumental terms, in view of our own goals—we clearly do. It just means that we can appreciate that others have value beyond their usefulness to us, and we are capable of respecting this, rather than always using other people merely for our own ends. I can love someone, for instance, who doesn't love me in return.

latter. Even being cognizant of this methodological guarantee, we may marvel at how narrow Kant manages to make the moral domain. For him, the sole criterion that determines whether something or someone is an end in itself or a mere means is the presence or absence of a rational capacity.[190] A being endowed with reason acts on the basis of principles that it gives to itself, and this freedom of will evidently sets rational persons outside of the causality that governs the rest of nature.[191] Since all non-rational nature is governed by causal laws, such nature is not free—thus, non-rational "things" have only a "relative" value, as means to the ends of rational beings.[192] Thus, Kant draws a line between, on one side, the rationality of rational beings, and on the other side, everything else in the universe. And he says, only the former is "good in itself," and so morality is ultimately about that. As a result of the narrowness of his ontological conception of intrinsic goodness, Kant's moral domain is strikingly narrow, with right and wrong being defined solely in view of our responsibilities to the "humanity" of others.[193]

Whatever one's views are in the domain of moral metaphysics, no psychologist can accept this definition. For us, there is an empirical side to the question of what determines the moral domain. It doesn't matter whether a philosopher would consider people's moral criteria to

---

[190] "The beings whose existence rests not on our will but on nature nevertheless have, if they are beings without reason, only a relative worth as means, and are called things; rational beings, by contrast, are called persons, because their nature already marks them out as ends in themselves, i.e., as something that may not be used merely as means." Kant, *Groundwork,* 46, AA 4:429.

[191] A freedom we must assume in order to act, Kant says, whether this assumption is correct or not. Kant, *Groundwork*, 64, AA 4:447-448.

[192] Kant, *Groundwork,* third section, 63-69, AA 4:445-453.

[193] Properly, our responsibility is to the rationality in any rational beings, whether those beings are people, angels, or rational aliens. This is "rationality" in a specifically Kantian sense—an ability to rationally conceive of the good and pursue it out of one's free will, rather than as a matter of causal necessity—and it essentially defines us, corresponding to the "humanity" in humans.

be universally valid, objective, rational, or any such thing. What matters is how people actually think. And when you adopt such an empirical approach to defining the moral domain, it is impossible not to see that morality is about much more than respect for other rational beings—we may also feel a sense of moral responsibility towards infants; non-human animals and plants; mountains, oceans, and earth; ideals and dreams for the future; even objects, insofar as they are sacred, sentimental, or beautiful. And our responsibilities towards these non-rational things arise, initially and for the most part, without tortured philosophical ratiocination. Instead, moral responsibility arises in view of any existential framing of a person, place, object, idea, or relationship as being intrinsically valuable[194]—whether because it is meaningful, beautiful, sacred, unique, precious, vulnerable, rational, or free. This normative interpretation of things (non-moral existential framing) organizes moral responsibility in advance of explicit moral reasoning, shifting the boundaries of our moral domain as we encounter different people, places, objects, and ideas, as our minds expand with empathy and experience or shrink with fear and prejudice, and so on. As such, while there are both universal and culturally specific trends for what humans consider to be intrinsically good and morally relevant, the actual domain of moral responsibility is ultimately determined by each individual in their unique context and moment.

This individual specificity of the moral domain may look a lot like moral relativism—the idea that right and wrong are subjective. Indeed, this definition has the plasticity to accommodate any moral metaphysical system, any personal moral philosophy, any transient moral intuition. But since I am proposing a merely descriptive model of psychology, my relativistic definition of

---

[194] "Intrinsic value" is how I am thematizing all the non-moral normative framing that is a prerequisite for moral framing. This is the derivation of the moral domain, which I promised you in chapter 1. I add the caveat, to be discussed in chapter 5, that the existential proximity of things also affects our sense of moral responsibility—as there may be valuable things whose well being we consider "not our problem."

morality does not actually involve any relativistic normative claim about right and wrong as such. Simply put, in order to be properly descriptive, my model *must* be able to describe how moral cognition happens, whenever and wherever and however it happens.

How is the moral domain delimited? In agreement with Kant, I acknowledge that the demarcation of the moral domain is initiated as we distinguish the instrumental value of things from non-instrumental value. Only non-instrumental framing (i.e., intrinsic goodness / badness) of things provides a basis for second-order moral framing.[195] Different moral philosophies propose different candidates for intrinsic goodness / badness: for Kant, the only thing that is good in itself is a rational will; for Bentham, intrinsic goodness and badness are reducible to pleasure and pain; Aristotelian ethics are conceived in view of the intrinsic good of Eudaimonia (a very expansive idea of "happiness"); Platonic metaphysics emphasize truth, beauty, and virtue; some philosophers have also built ethical philosophies in view of intrinsic badness, such as the ethical responsibilities that arise through our encounters with suffering (e.g., Emmanuel Levinas). My polyphonic model makes no claim as to which of these systems is best but offers a principle that is common to all of them and more. All moral normativity is ultimately about things that have already been deemed intrinsically good or bad. Thus, we can phenomenologically complete the operational definition of morality from chapter 1. Morality is a second-order normativity concerned with how we ought to perceive, judge, desire, act, and be, in

---

[195] This is what Kant would call an "analytic" proposition, because I am merely elucidating meaning that is already there implicitly. Instrumental things are, by definition, means to some other ends, and their value as such is always relative to the ends they serve, which ends themselves may be judged good or bad. We can only have moral responsibilities towards things we have already deemed intrinsically good or bad. We should recognize, however, that things that have instrumental value may *also* take on intrinsic value—e.g., the aesthetic and sentimental value of a classic car. Thus, when I am talking about instrumental things, I am speaking more precisely of their *instrumentality*—the thing *as* instrument, and not the thing *as whatever else* it may be.

view of whatever is framed as intrinsically good or bad, whatever matters to us in a way that is not merely instrumental.[196] Every moral system, principle, or sentiment is grounded on an obligation to "take care of what is good / oppose what is intrinsically bad," an imperative that acknowledges that one's understanding of what is intrinsically good or bad varies among cultures, philosophies, individuals, and moments.

How does this deeper understanding of the meaning of morality speak to current moral psychological theory? Consider the ongoing debate about the distinction between moral rules and rules of convention (recall that psychologists describe moral rules as relatively universal and context independent, while conventions are seen as context-dependent and situational). Some psychologists claim that moral rules are always injunctions against harm, while others argue that what counts as a moral rule varies from culture to culture and involves considerations beyond harm. In support of the latter pluralist claim, researchers adduce acts that cause no evident harm to anyone (such as privately using a flag to clean a toilet or engaging in harmless but aberrant sexual acts), but which some populations consider to be wrong, independent of context.[197] Researchers on the other side of this debate adduce studies showing that many people

---

[196] It also matters just how valuable intrinsically valuable things are, and just how deliberate our actions towards these things are thought to be. Many people who consider life intrinsically valuable would nonetheless stop short of describing the killing of an ant as "immoral," even if this killing were done for no good reason. Presumably, such people view the ant's life as too trivial to warrant invoking the idea of morality. Some, however, might describe the senseless killing of an ant as an immoral act. There is also a spectrum of intentionality that factors into our sense of whether to invoke the idea of morality. We would be more likely to judge the killing of the ant as immoral the more deliberate we consider the killing to be—if, say, someone killed the ant on purpose, just to kill it and for no other reason. In invoking the idea of morality, we are thus sensitive to a subjective threshold of perceived deliberateness and value-magnitude, which must be crossed before we will feel justified in invoking the idea of morality. I will call this the "morality threshold." My definition of morality presumes that the morality threshold has been reached.

[197] Haidt, "Is it Wrong to Eat Your Dog?" 613-628.

automatically import an idea of harm into these scenarios.[198] The pluralist response is that the idea of harm that people import cannot completely explain all moral objections in these cases, let alone all possible cases.[199]

The foregoing analysis suggests that harm is a bad criterion for delimiting the moral domain, not because of a pluralist critique of moral monism but because, regardless of how many moral values there are, the very idea of a "moral" value presupposes an idea of morality the conditions for which must be articulated. Moral values are secondary moral forms, renegotiations of a moral sensibility that arises initially through existential framing, in response to the intrinsic value we attribute to beings in the world. We know that morality depends upon attributions of intrinsic value, as I indicated above, because the idea of moral obligations in the absence of intrinsically good / bad things is absurd. Moreover, the idea of intrinsic value helps us resolve the above argument between the monists, who emphasize the universality of morality, and the pluralists, who emphasize that different cultures have different moral values. Intrinsic value is both universal and relative. It is universal in that, if the value of something is understood to be intrinsic to the thing, this value is thus *understood* as being independent of any cultural or subjective opinion. And the moral obligations that arise from such value attributions are accordingly understood as being universal in the same sense. Nevertheless, from a descriptive psychological perspective, we must acknowledge that what one views as intrinsically valuable—

---

[198] Chelsea Schein & Kurt Gray, "The Unifying Moral Dyad," 1147–1163; Edward Royzman, et al., "The Curious Tale of Julie and Mark," 296-313.

[199] Jesse Graham et al., "Moral Foundations Theory: On the Advantages of Moral Pluralism Over Moral Monism," in Kurt Gray & J Graham (Eds.), *The Atlas of Moral Psychology: Mapping Good and Evil in the Mind,* eds. Kurt Gray & Jesse Graham (New York: Guilford, in press); Joshua Rottman, et al., "Purity Matters More than Harm in Moral Judgments of Suicide," *Cognition* 133.1 (2014): 332-334.

and how valuable one considers it to be—varies from culture to culture, person to person, day to day, even mood to mood. And accordingly, from a descriptive perspective, we must acknowledge that morality is relative.

Just as to define physical motion, one must begin by identifying one's frame of reference; to define the morality of an act, we need to begin by identifying which existential frame we are in. Most crucially, we must locate the frame(s) of intrinsic value. If different people have different frames of intrinsic value, their moral perception will be different, just as people in different inertial frames perceive motion differently.[200] For some people, a flag is understood simply as a symbol, the value of which is entirely relative to its expression in a social context. A particular flag is not intrinsically valuable, but is merely a means to a valuable end—symbolic expression, the value of which is also relative to other intrinsically valuable endeavors. As such, the morality of one's treatment of the flag is determined by the social context of the flag's use; in private, a flag may be used as material for other projects, even toilet cleaning. For others, however, the material flag is itself sacred. As such, it is intrinsically valuable, and even private treatment of the flag will be governed by moral law. Thus, in this case (as in others) the distinction between conventional and moral violations can be explained not by referring to a moral value like purity or care for others—which begs the question of what makes these values "moral"—but instead by identifying the existential frame within which the flag is viewed,

---

[200] I am thinking now of Einstein's example, at the beginning of *Relativity,* of the motion of a rock dropped from the window of a moving train. From the perspective of the person on the train, the rock falls in a straight line. Relative to someone standing on the tracks, the rock performs a parabolic curve. Neither pattern of motion is objectively true. One's description of the motion depends on one's frame of reference (Albert Einstein, *Relativity: The Special and the General Theory*, trans. Robert W. Lawson [New York: Barnes & Noble, 2008 (original 1920)], 9-10).

recognizing that what is seen as intrinsically valuable may vary from culture to culture, person to person, and situation to situation.

Polyphonic Consequences of Existential Framing: The Officer's Dilemma

So, the entire domain of morality is demarcated in view of this distinction between instrumental and non-instrumental existential framing—between things that are valued merely as means to other ends and whatever is valued for its own sake. Seems simple enough, right? But this conceptual simplification does not simplify moral cognition in practice, because, as we have already seen, many things are viewed within multiple existential frames, multiple normative modalities. An object may be both useful and beautiful, for instance. Even if we follow Kant and restrict ourselves to the human domain, we must acknowledge that it is impossible to wall people off from all instrumental and calculative modes of perception. While I may recognize and respect others' autonomy, this does not mean that I stop judging them in instrumental ways. It still matters whether they might be effective colleagues, good lovers, fun companions, valuable allies, and so on—that is, even as I respect others as ends in themselves, I continue to evaluate them in view of instrumental ends, my own or otherwise. And this complicates my moral life greatly.

Let's consider an example. A commanding officer relates to the members of their division as both soldiers and human beings. In their capacity as soldiers, members of the unit are defined by the role they play, and they are judged as good or bad soldiers in view of how they serve the goals of the unit, contextualized by the intrinsic value of other members of the unit, as well as the good of the country / citizens they serve. But these same individuals are also women and men, who are ends in themselves, whose lives matter independent of any cost-benefit analysis their superiors might make. As a result, commanding officers may face difficult moral

dilemmas that are brought about by this essential conflict of existential framing, as they balance what is strategically best against what is best for individuals in their unit. Such dilemmas may appear to express a conflict between "utilitarian" and "deontological" rationalities—with utilitarian logic emphasizing the ultimate benefit for the greatest number of people, and perhaps privileging the strategic good, and deontological logic forbidding any reduction of a person for use as a mere means to an end. But such moral conflict can also be understood more simply, as a conflict between what is good in two distinct existential frames. Insofar as the commanding officer approaches the individual as a soldier—defined in view of a functional role within the unit—the officer's understanding of what's best for the unit (in view of the intrinsic value both of other members of the unit and the intrinsic value of the country / citizens the unit serves) is the moral priority. Insofar as the officer approaches the individual as a person who is good in his or herself, the officer feels morally obligated towards the person. Moral conflict arises simply because moral norms in the frame of soldierhood may conflict with moral norms in the frame of personhood—distinct frames through which the commanding officer views the same individuals. The officer doesn't need to refer to a principle of maximal utility or a categorical imperative to experience this moral conflict. The possibility of such conflict arises simply from an appreciation of the dual ontological status of these men and women, simply from an appreciation that these are both soldiers and human beings.

Even in the absence of explicit moral conflict, we are constantly inhabiting this moral polyphony, simply by encountering and interpreting people and other beings in multiple existential frames. And this polyphonic reality would be true, as indicated above, even for a "monistic" theory of moral psychology. Even if we only consider harm, for instance, there remains a fundamental difference between the commander's moral responsibilities towards the

soldier and their moral responsibilities towards the person. What is best for protecting the unit, country, or citizens from harm may conflict with what is best for protecting an individual in the unit. But polyphony pervades much more than such situations of moral conflict. Moral dilemmas are only the most visible and dissonant form that this polyphonic reality takes. Whenever the good of the soldier harmonizes with the good of the person—as it often does—we may forget this underlying multiplicity of our relations. But this forgetting changes nothing. The reality is that we constantly encounter people, places, objects, ideas, and so on within multiple existential frames—each of which relates to distinct intrinsic goods / bads—and as a result, polyphony is an inescapable reality of our moral-psychological condition.

Existential Framing of the Self

In chapter 2, I argued that psychologists need to adopt a thoroughly biological view of the psychological processes that underlie moral judgment. If we understand that our psychology has been built "from the ground up," we are able to give up the expectation for our psychological development to proceed in a linear way towards a single unambiguous good—and we are prepared to appreciate the real polyphonic dynamics underlying moral cognition. However, we must also avoid the pitfall of trying to understand ourselves purely as biological organisms. Our existential identity extends beyond our biological bodies. I am not only a 6 foot 5 inch, 190 pound, blond-haired, blue-eyed male, with such and such cognitive capacities, and so on. More important from a moral-psychological perspective is the fact that I am my mother's son, my sister's brother, and a citizen of America. More important is the fact that I have a unique past, parts of which I want to honor and parts of which I would rather erase, that I have ambitions and hopes for my future, that I am engaged in projects that are meaningful to me. These facts about

who I am cannot be predicted from my biology, except in the vague sense that I am the kind of organism that is designed to form such meaningful attachments and involvements. That is, a certain biological makeup is a condition for the formation of this identity; nevertheless, the identity that forms through these attachments is extra-biological and unique. The self that is the subject of moral judgment is largely distributed *out there* in these existentially meaningful nodes, in my attachments to specific people, places, nostalgic memories, hopes, ideals, and so on.

Another way of saying this is that the self is not an objective thing that *has* a mind but, rather, a subjective reality that the mind creates. As such, we might speak of "self-creation" as something that the human organism *does*. As the mind forms relationships, and as it tells stories about those relationships, it builds and maintains a certain self, a network of nodes of selfhood, a "selfscape." In striving to maintain our moral identity, we are maintaining the relationships that form this broader selfscape of our moral identity—relationships that are external to our biology but internal to our self. To *be* a commander or friend or soldier or son is to tell oneself a certain story, a story that informs not only what or who one understands oneself to be but also how one thinks one *ought* to be.[201] Self-creation thus involves constructing existential frames. I can understand whether my behavior is good or bad with reference to the existential frames of sonhood or soldierhood—which include ideas of how a son or soldier ought to behave in various situations. And when my actions fail to live up to my idea of a good son or soldier, I haven't just

[201] Sociological theorists like Robert Bellah (*Habits of the Heart*), Christian Smith (*Moral, Believing Animals*), and Charles Taylor (*Modern Social Imaginaries*), have offered illuminating characterizations of some of the powerful stories that function within cultures and subcultures to create shared identities and support certain moral orders. I acknowledge that such sociological models are accounts of a certain kind of existential framing—of identities that are constructed and shared by a social group—but I am interested in *all* forms of existential framing of individuals, groups, places, things, and ideas, from the most transient and idiosyncratic (the framing of the moment, Raskolnikov's idea that "nobody had ever thought before") to the most universal.

done a bad thing, I've actually *been* a bad son or a bad soldier—I've suffered an existential wound to my self. The contours of my moral identity are thus determined not only by my internal capacities, beliefs, and dispositional qualities but also by my external attachments and pursuits. Thus, we see that there is a certain commonality to parochialism and impartiality, beyond the claim I made earlier that both are grounded in Sorge: just as my parochial attachments are internal to my existential identity, so are my ideals of impartiality, justice, and universal rights. Insofar as they are *my* ideals, which I identify with, they are also internal to my self. In the same way that I *am* my friends' friend, I *am* impartial and just. And just as I am involved in selfish projects, I am also involved in "selfless" endeavors—selflessness is, in this sense, a particular mode of self.

So, existential framing is not just something that we impose upon other objects or people but also an interpretive framework through which we understand our*selves*. I *am* my mother's son, my friends' friend, my country's citizen, my cat's owner, and so on. I am a person with certain ideas, who does certain work, who belongs to a certain political party, who is working with some sense of purpose towards certain goals. My relations to people, places, ideas, non-human animals and plants, meaningful objects, and so on define not only what or who these external things are but also what and who I *myself* am. And again, these ontological relations have immediate normative consequences. *As* a friend, I implicitly understand myself to have certain obligations towards my friends. I have obligations towards my pet as an owner, obligations towards my planet as an inhabitant or steward, obligations to behave in accordance with my core beliefs, obligations towards my family, my country, my teammates, and so on. The intimacies and loyalties that connect me to other people, places, creatures, and objects, the convictions that connect me to moral and political beliefs, simultaneously define who / what I am

and dictate how I ought to be—from how I should behave to how I should perceive, judge, and even feel in diverse situations.

This understanding of selfhood in terms of the meaningful attachments one forms and the projects one has, along with the insight that this distributed selfhood determines much of moral perception, judgment, and action, is missing from theories that attend only to nature and nurture, biology and culture, and the interaction of the two. Moral psychologists tend to invoke biology to identify a generic biological or developmental profile that holds across populations, and they tend to invoke cultural learning when pointing out moral differences between cultural groups. Thus, bio-cultural interaction is not used to explain the unique morality of a particular individual at a particular moment but, rather, to describe comparatively stable group-level traits. For the specificity I am seeking, we need to invoke existential framing. Existential framing explains the *uniqueness* of who one is in a given situation, and shows how this unique interaction of an individual with the broader environment gives rise to a moral sense that fluctuates from year to year, day to day, even mood to mood. What we must understand is that the unique relationships that constitute selfhood and frame our moral perception, judgment, and action do not constitute mere noise, which a good psychological theory should smooth over. Rather, such parochialism, "irrationality," and idiosyncrasy of our moral psyches has a perceivable logic, which is universal for all moral beings. Our loyalties, our sense of what is just, our capacity for compassion, our understanding of legitimate authority, and on and on, are largely determined in view of the involvements that constitute who we are as unique individuals in a given moment. And, as we have seen, it is precisely the ability / tendency to form meaningful involvements that is *universally* decisive for moral cognition and behavior.

Conclusion on Existential Framing

The preceding analysis of existential framing exposes several facts that I regard as undeniable, but which have nevertheless gone unappreciated in the field of moral psychology. I began by asserting that the unique and the universal, the parochial and the impartial have a common potentiating source. I then demonstrated that moral judgment is always being determined in view of prior non-moral evaluations of things. This extra-moral attribution of intrinsic goodness / badness creates a moral topography that is specific to each culture, individual, moment, and mood. And yet, because everyone has an understanding that is structured by Sorge, this dynamic of existential framing is itself universal. Finally, I argued that, because people, places, creatures, objects, ideas, and relationships are constantly viewed within multiple existential frames, multiple evaluative modes, we must constantly negotiate a plurality of imperfectly aligned moral responsibilities towards all these intrinsically valuable things—and thus, moral cognition is pervasively polyphonic.

It should also be clear that moral polyphony is not another name for Haidtian value-pluralism. Polyphony arises through diverse existential involvements that exist prior to all morality and, a fortiori, prior to moral values: thus, even a "monistic" theory that recognizes only one moral value may be (should be) polyphonic. There is much more to be said about existential framing and its importance for moral psychology. But I am ready to apply the above theory to another reading of Raskolnikov's confession. I include further theoretical discussion in endnotes: Mediating Factors in Existential Framing: Systemic Relations Among Objects and Concepts[i], Politics and Existential Framing[ii], Existential Framing and Empathy[iii] Existential Framing and Heidegger[iv], and Equivocality of the "Existential Framing" Concept[v].

Raskolnikov's Confession Revisited

"I only killed a louse, Sonya, a useless, nasty, pernicious louse." "A human being—a louse!" "Not a louse, I know it myself," he replied, looking at her strangely. "Anyway, I'm lying, Sonya," he added, "I've been lying for a long time…"[202]

Why Confess?

When we first encountered Raskolnikov's confession to Sonya, back in chapter 1, I asked you to bear in mind the question why Raskolnikov feels compelled to confess at all. Now I invite you to consider how strange it is that anyone should ever want to confess wrongdoing. This everyday phenomenon cries out for an explanation, especially in Raskolnikov's case. He is not repentant. He is not convinced that what he has done is wrong. He can be sure that, even if he manages to convince Sonya of the moral defensibility of his crime, she will not approve of it. Given that he evidently stands to gain nothing from a confession, but actually risks his relationship and potentially his freedom (should Sonya decide to turn him in), why is he so driven to confess? Why isn't he content, knowing that he's gotten away with the crime, and that Sonya will never suspect him of it?

Do existing moral-psychological theories have a good answer? We find the beginnings of one in Adam Smith's *Theory of Moral Sentiments*, of 1759. Smith proposes that moral judgment is undertaken by perceiving the actions of others and oneself from the point of view of an "impartial spectator."[203] We are able to imagine how we appear to others and to judge our own actions and character using the same standards we apply to other people. Thus, Smith says, if our minds are "well-formed," we cannot be content with our own wrongdoing, even if no one else

---

[202] Dostoevsky, *Crime*, 416.

[203] Adam Smith, *The Theory of Moral Sentiments* (Oxford: Oxford University Press, 1976 [Originally published in 1759]), 93.

knows about it—we desire not merely to be praised but also to be praiseworthy.[204] This desire could partly explain why Raskolnikov wants Sonya to know the truth. It obviously matters to him not just that he *appear* innocent to Sonya and those close to him, but that his action might actually be vindicated. Raskolnikov is not a petty sociopath, like the capitalist Luzhin; and it is not in his nature to be content with Sonya's esteem for him, if that esteem is based on a lie. Nevertheless, even accepting this view of Raskolnikov's psyche, we must admit that it does not really explain why he should want to confess. He could, after all, rationalize his action privately, without feeling compelled to externalize this justificatory process, and he has indeed been engaged in such rationalization, since well before the crime.[205]

We might get closer to an answer with Haidt's Social Intuitionist Model. We are social beings, and moral reasoning is socially motivated, Haidt might say, so it might really matter for Raskolnikov to justify himself to Sonya. Perhaps he feels that if he can convince Sonya that his action was morally defensible, he will be able to convince himself too, and his guilt will abate. However, we must acknowledge that this is still at best a partial explanation. While he initially seems to want to justify himself to Sonya, to convince her of the moral defensibility of his crime, Raskolnikov later pursues the opposite tack, provoking Sonya to see his action in an unfavorable light and portraying himself by turns as a petty thief,[206] a narcissist,[207] and an incipient

---

[204] Ibid., 96-99.

[205] Dostoevsky, *Crime,* 70: "it would seem he had already concluded the whole analysis, in terms of a moral resolution of the question: his casuistry was sharp as a razor, and he no longer found any conscious objections."

[206] Dostoevsky, *Crime*, 412.

[207] Ibid., 417.

madman.[208] Thus, just as surely as Raskolnikov's confession aims at a justification of the crime, it also aims at times at a condemnation of the criminal, and this combination is something that neither Smith's account nor Haidt's is designed to explain. Clearly, there is something else driving Raskolnikov to confess, beyond his need to justify himself to himself and beyond the social nature of this justificatory process.

In fact, Raskolnikov himself seems not to understand why he is compelled to confess to Sonya. He expresses this confusion three times: first, asking himself why he has come "to torment" her;[209] then, after he has just confessed, saying to himself, "why did I tell her, why did I reveal it to her?"[210] and finally, asking rhetorically again, "why did I come?"[211] He offers several unsolicited answers. He came to ask for "forgiveness."[212] He came to ask her not to leave him.[213] He came to "shift the burden" onto her.[214] We have already considered the first of his answers. Appreciating that the "forgiveness" Raskolnikov was pursuing involved a justification of the action and not an expression of remorse, I concluded that the pursuit of such "forgiveness" is at best a partial explanation for his need to confess, since he then goes on to condemn himself (without thereby expressing an attitude of repentance). The second answer does not really pertain to the confession at all. Sonya is not about to leave Raskolnikov, and he probably knows that she

---

[208] Ibid.

[209] Dostoevsky, *Crime,* 409.

[210] Ibid., 414.

[211] Ibid.

[212] Ibid., 408.

[213] Ibid., 414.

[214] Ibid.

will stay with him whether he confesses or not. To ask for her companionship is one of the reasons he came—it is why he had "called" her to him the day before[215]—but it does not explain why he is compelled to confess. So, we are left to interpret the third answer.

What is the "burden" that Raskolnikov wants to share? Over and over, in the aftermath of the crime, the presence of those Raskolnikov loves, and especially any kindness and concern they express for him, is portrayed as being a torment to him. We see this from early in the novel, with his responses to his mother and sister. Even though he has been longing to see them after three years' separation, when they finally arrive in Petersburg he is overwhelmed:

> A cry of rapturous joy greeted Raskolnikov's appearance. Both women rushed to him. But he stood like a dead man; a sudden, unbearable awareness struck him like a thunderbolt. And his arms would not rise to embrace them; they could not. His mother and sister hugged him tightly, kissed him, laughed, wept … He took a step, swayed, and collapsed on the floor in a faint.[216]

Raskolnikov does not immediately understand why seeing his family is so difficult, musing about how he "seemed to love them so much when they weren't here."[217] But presently it dawns on him that it's because of the crime. This realization happens in their first substantive conversation after his fainting spell. He cuts the painful conversation short by telling his mother that they will "have time to talk all they want!"[218] But immediately it becomes "perfectly plain and clear" that it is "no longer possible for him to talk at all, with anyone, about anything,

---

[215] Ibid.

[216] Ibid., 192.

[217] Ibid., 227.

[218] Ibid., 229.

ever."[219] In the aftermath of the murder, "this familial tone of conversation, together with the complete impossibility of talking about anything at all" is almost unbearable to him.[220]

So, the burden is Raskolnikov's isolation in the wake of the crime. To share the burden is to be able to share at all. And this need to connect with Sonya is a major part of the answer to the question why Raskolnikov is so driven to confess. Why is this? Why doesn't Raskolnikov simply lie about his crime and thereby protect his relationships, while speaking candidly about other things? Well, he does lie. The real question is—why do his lies fail to protect him from misery? The answer is that, as Adam Smith might point out, even though his loved ones do not know he is lying, Raskolnikov knows, and this awareness is itself painful. Contra Smith, however, we must acknowledge that this torment exists for Raskolnikov even though he is not convinced his action is wrong. Indeed, Raskolnikov approves of the crimes of people he views as exceptional, and he considers it their duty to "step over" all obstacles in pursuing their vision of the future. Thus, it is not Smith's "impartial spectator" who torments Raskolnikov, but something altogether different.

In chapter 1, I pointed out that Sonya seems to understand something Raskolnikov does not, in spite of all his moral ratiocinations. This is most explicit in the aftermath of his confession, when a broken Raskolnikov asks Sonya what he should do. She has a strikingly specific answer ready: "Go now, this minute, stand in the crossroads, bow down, and first kiss the earth you've defiled, then bow to the whole world, on all four sides, and say aloud to everyone: 'I have killed!' Then God will send you life again."[221] I indicated then that this model

---

[219] Ibid.

[220] Ibid., 231-232.

[221] Ibid., 420.

of repentance is expressed in all of Dostoevsky's major novels, and I pointed out that this form of moral resolution involves a rejection of the possibility of achieving inner peace through philosophical ratiocination—a project that, ironically, can only succeed through an irrational rejection of the reality of moral polyphony—and advocacy of a return instead to innocence and faith, Dostoevsky's characteristic antidotes to nihilism.

Sonya's answer to Raskolnikov suggests a deeper question. Even if we accept that all the ratiocination in the world can never give Raskolnikov anything more than an unstable, illusory moral coherence, we may well ask why repentance should work any better. Sonya seems to know why. Although she does not express her answer in the form of a theory, implicit in her responses to Raskolnikov is an understanding that his betrayal of external relationships is at the same time a self-wound. For instance, after a brief spell of terror, Sonya's initial reaction to Raskolnikov's confession is an outpouring of pity, not for the victims but for Raskolnikov himself: "'What, what have you done to yourself!' she said desperately, and, jumping up from her knees, threw herself on his neck, embraced him, and pressed him very, very tightly in her arms."[222] Sonya's pity presupposes that, in murdering Alyona Ivanovna and Lizaveta, Raskolnikov has not only done external violence but has also inflicted great *internal* damage—that is, this pity only makes sense in view of a distributed model of moral selfhood, such as the one I outlined above, where one's external relationships are internal to one's self. Such a distributed model of the self also explains Sonya's insistence that Raskolnikov repent. If his suffering stems from the fact that his external act of violence also created an existential wound

---

[222] Ibid., 411-412.

within himself, it follows logically that internal psychological healing will involve repairing the external relationships.

There is, of course, no way to make amends to the dead, nor perhaps to the society whose rules Raskolnikov has violated. But existential repair of the self is fundamentally a matter of repairing one's *relations* to others. And simply taking steps to realign one's attitude and intentions can facilitate such an existential repair, even when it is impossible to make amends. As long as Raskolnikov rejects the fundamental values of his society and refuses to repent of his crime, as Sonya sees it, he precludes any repair of his internal schisms. But if he genuinely repents, even though he can never make amends for his actions, he can create the conditions for his bonds to God, the earth, society, and those he loves to be renewed and his identity to be restored. Thus, when Raskolnikov rejects Sonya's advice and says he won't turn himself in, she insists that this path is totally untenable for him, precisely because of what it means for his relations to others and what these relationships mean for his own existence: "'how will you live? What will you live with?' Sonya exclaimed. 'Is it possible now? How will you talk to your mother? (Oh, and them, what will become of them now!) But what am I saying! You've already abandoned your mother and sister. You have, you've already abandoned them.'"[223]

In his principled justification of murder, Raskolnikov has split from his family, his friends, his city and country and religion. Sonya's insight is that this is not merely an external split but an internal wound. At some level, Raskolnikov is aware of this too. He imagines that he would be happy if he killed merely because he was hungry[224] and not because he believed that he had the right to kill. If he killed out of desperation and felt sorry about it, he would at least

[223] Ibid., 420.

[224] Ibid., 413-414.

maintain his connection to the moral ideals of his friends, family, and country. His relationships would be damaged, but not beyond repair. He would feel sorry and repent, and the connection to these others would begin to heal—indeed, as long as he was genuinely repentant, he would never be entirely cut off from these relationships. But Raskolnikov refuses to repent. And this refusal, along with Sonya's own refusal to accept early versions of his account, is what draws the deeply polyphonic dynamics of Raskolnikov's psyche up to the surface of his confession. Rather than simply expressing remorse for an unjustifiable but understandable misdeed—and thereby leaving his true psychic complexity safely submerged—Raskolnikov is compelled to dredge up all the competing loyalties that factored into his decision.

Beyond the fact that the central drama of *Crime and Punishment* presupposes a model of distributed selfhood, I believe that implicit in Raskolnikov's confession is a corollary suggestion about repentance, a suggestion that is in keeping with the spirit of Dostoevsky's Christianity, but which might nonetheless seem heretical to some believers. *The need for repentance is independent of the morality of one's actions*. It is a psychological need that is totally independent of any metaphysical moral truth. Even though Raskolnikov considers his action morally justifiable, he still has a psychological need to repent and thereby renew his relations with others. This need arises from the existential rift between Raskolnikov and his friends, family, and society, which he experiences as a burden or wound. Justifying his action to himself does not heal or disburden Raskolnikov, because it does not heal the underlying existential wound. In fact, his philosophical ruminations seem to make things worse, as he must deal not only with the angst of the existential rift he has created with the murder, but also with resentment aching from another existential rift, between himself and a society that supports so much daily injustice. Indeed, this resentment is at the root of Raskolnikov's initial desire to commit the crime, since he

conceived the murder in part as a rejection of the authority of an absurd social order.[225] The resentment marking Raskolnikov's estrangement from his society long preceded the crime. And the same thoughts that whispered justifications for murder back then now foment continued anger against the society whose laws he has broken, a resentment that renders the very idea of repentance intolerable, as we see in his response to Sonya's final plea:

> "But how, how can one live with no human being! What will become of you now!" "Don't be a child, Sonya," he said softly. "How am I guilty before them? Why should I go? What should I tell them? It's all just a phantom … They expend people by the million themselves, and what's more they consider it a virtue. They're cheats and scoundrels, Sonya!… I won't go."[226]

Thus, Raskolnikov's need to confess arises from the fact that, even though he does not acknowledge the moral wrongness of his action, his action nonetheless damages his relations to those he loves and the society of which he is a part, and this damage is internal to his existential identity. Repentance would be a step toward healing this existential wound, but Raskolnikov's resentment toward society stands in the way. For the time being, all he can do is confess to Sonya, in a dialogue that expresses the intractable contradiction between his desire to justify his action and his need to be forgiven.

Raskolnikov's Identity is *Not* Merely a Social Identity

The above reading emphasizes the social dimension of the self, and readers may want to point out that the idea that humans identify very much with our social groups, and that social identities shape our moral worlds, is not new. Contributions to moral psychology from

---

[225] Ibid., 418.

[226] Ibid., 420-421.

sociologists, anthropologists, social psychologists, and even evolutionary theorists often emphasize the hyper-sociality of human nature. For instance, in their celebrated defense of the role of group selection in shaping human altruism, *Unto Others*, Elliot Sober and David Sloan Wilson emphasize the many ways in which humans think and act as members of social groups and not merely as selfish individuals—to an extent that far exceeds the altruistic and cooperative tendencies of other social primates. Haidt identifies this pervasive human "groupishness" as the source of sacred values and, building both on Sober and Wilson's evolutionary account and on Emile Durkheim's sociological model, he emphasizes the power of sacred ideals to bind human groups together.[227] Such accounts are very much in line with what I have proposed, but I offer a phenomenological complement to these theories and expand the horizon of moral relevance to include all people, places, creatures, objects, ideas, and relationships—all beings, in short—that are understood as being intrinsically valuable. And I maintain that, although there are commonalities within cultural groups, ultimately the moral domain is uniquely determined by each person, situation, and mood. Thus, even acknowledging that Sonya's recommendation for Raskolnikov may be all about repairing his relations to other people—and thereby repairing the self—we should continue to bear in mind that morality is more than a subdomain of human sociality, and we should continue to rely on the more expansive concept of existential framing to interpret his confession.

Consider, in this connection, Raskolnikov's most idiosyncratic reason for committing the crime—he "wanted to become a Napoleon,"[228] an exceptional type of person. Here, Raskolnikov

---

[227] Jonathan Haidt, *The Righteous Mind* (New York: Pantheon. Knopf Doubleday Publishing Group, 2012), Part III.

[228] Dostoevsky, *Crime,* 415.

seeks to establish an identity that is defined by its *difference* from the rest of the social group. And this extra-social identity offers a justification for crime. In an early discussion with the detective Porfiry Petrovitch, Raskolnikov explains his idea that an exceptional type of person, one who is capable of speaking a "*new word*," has a right, even a duty, to break all barriers and pursue their unique vision of the future.[229] This vision is thus a part of their identity, an intrinsic good that exceptional types are morally responsible for.[230] Because Raskolnikov's idea is opposed to the values of his social group, and he must therefore choose which intrinsic good to prioritize, there is actually no perfect solution that will offer him moral peace-of-mind. If he kills and steals, he violates his social self; but if he stops short, he abandons his idea. Sonya's suggestion that he repent and renew his bonds with society thus requires Raskolnikov to repudiate his highest ambition, a repudiation that would also involve an existential wound. Indeed, his existential investment in his idea is so profound that by the end of the confession it is evident that Raskolnikov is most ashamed not of the crime he has committed but, rather, of his inability to do it the right way, to transcend his guilt, to embody the subjectivity he attributes to the exceptional type of person.[231] Thus, it is evident that Raskolnikov feels a kind of moral imperative *not* to repent, not to denigrate his idea, just as he feels moral pressure to repent and renew his bonds with his society. There is no straightforwardly "right" course of action for him, because his moral world is polyphonic. And, crucially, this polyphony of his moral identity extends beyond the boundaries of his social world.

---

[229] Dostoevsky, *Crime,* 260: "it is even their duty."

[230] Recall that I've defined the moral domain as the normativity of perception, judgment, and action, in view of whatever is framed as being intrinsically good or bad. Other definitions of the moral domain might consider this particular responsibility to be non-moral—but I have already pointed out what I consider the absence of any principled reason for such an exclusion.

[231] Ibid, 419-420.

Raskolnikov's idea is just one example among the many extra-social things that people may regard as intrinsically valuable. We may care deeply for familiar places; sacred objects; animals and plants; earth, ocean, and sky; stars and planets; ideas and ideals; hopes and dreams. As such, we may develop a sense of responsibility towards these beings for their own sakes. Theories that try to explain morality simply in terms of our responsibilities as members of social groups offer imperfect accounting of moral cognition and action towards such extra-social things. Typically, such theories seek to explain extra-social moral normativity either in view of some cultural or evolutionary utility at the level of social groups, or as a side-effect of biological and/or cultural evolution. Even if such accounts are perfectly accurate, we must respect the distinction between different levels of analysis, such as the difference between "ultimate" and "proximate" analyses elaborated upon in chapter 2. An evolutionary, or even a sociological account cannot stand in for a model of psychological or phenomenological dynamics, and deferrals to these other kinds of analyses may simply hide the fact that such theories lack a principled way of delimiting the moral domain at the proximate analytic level of cognition. By contrast, I see no principled basis for excluding extra-social involvements from theories of moral cognition, or for theoretically reducing non-social attachments to some purely social meaning. And I maintain that the moral importance of our meaningful involvements with non-human things is obvious and undeniable. Humanity is not the only thing which people may consider intrinsically valuable, or "good in itself." And it doesn't matter whether Kant or anyone else thinks this is rational or not, because our task as moral psychologists is not to defend a rational conception of the good but to describe moral cognition as it happens to happen.

Conclusion: Existential Framing and Moral Cognition in Real Time

Although existential framing is the most distinctive and important feature of my model, it is not intended to replace existing research on socio-moral cognition. My analysis shows that existential framing is a transcendental condition of moral cognition, but we still need to trace the various faculties through which moral cognition happens in the real world. Happily, all the cognitive abilities and tendencies outlined in the first part of this chapter are on vivid display in Raskolnikov's confession. We see contradictions between different moral laws applied to the same situation. ToM computations are performed on virtually every page, most dramatically in the initial wordlessness of the confession itself; and we witness the power of the empathic connection between Sonya and Raskolnikov, as thoughts and feelings seem to pass contagiously between them. Character is a central focus, and Sonya's refusal to accept early iterations of the confession owes to the fact that the motives Raskolnikov offers as explanations for his crime are incongruent with the idea she has already formed of his character. Finally, we see how feelings of compassion, resentment, and suffering not only attend Sonya's and Raskolnikov's thoughts but also play a central role in the cognitive dynamics of the confession, provoking not only affective, emotional, and locutionary responses, but also shifts in existential framing itself. Let's look at these cognitive features one at a time, using the theory from the first section of this chapter to interpret the cognitive dynamics expressed in this exchange.

I begin with moral rules. Raskolnikov and Sonya have different ways of making sense of the murder. Whereas he rationalizes his action in view of several distinct moral criteria, including both consequentialist and virtue ethical principles (albeit of a uniquely Raskolnikovian form), Sonya acknowledges the authority only of God's moral law (a conviction that, in Greene's usage of the term, would be called "deontological"), and so she seeks to understand at a more intuitive level what must have transpired in Raskolnikov's psyche or "soul" to make him do such

a terrible thing. Not understanding Raskolnikov's enigmatic claim that he killed because he "wanted to become a Napoleon," Sonya entreats him to continue: "just go on! I'll understand, I'll understand everything within myself!"[232] Unwilling to accept rationalizations that seem to break too starkly with her understanding of who Raskolnikov is—despite the fact that these accounts are sincerely spoken[233]—Sonya nonetheless trusts that as she listens to his explanations she will intuitively grasp the truth within herself. This truth is not contained in any of the consequentialist rationales Raskolnikov offers, nor in his ambition to become a higher kind of person—though these rationales do also speak within Raskolnikov's polyphonic psyche. The truth that Sonya ultimately takes away—that Raskolnikov has rejected God and God's laws and been confounded by the devil[234]—is a more hopeful truth, because it preserves the core of her theory of Raskolnikov's character, and it offers the possibility that he might be restored through repentance.

Sonya's overwhelming compassion for Raskolnikov is surely enabled by the fact that she already loves, respects, and trusts him. Otherwise, her caring response to him wouldn't make any sense, given that he has just confessed to having murdered one of her only friends. If she thought he was a sociopath like Luzhin, she would surely not have rejected his early claim that he killed simply to rob, and she would have remained in the state of terror that accompanied her first recognition of what Raskolnokov was trying to confess. Instead, we see her struggling to reconcile Raskolnikov's crime with the person she knows and loves—a difficulty that leaves her,

---

[232] Ibid., 415.

[233] "Ibid., 416. "You can see for yourself that's not it!… yet it's the truth, I told it sincerely!"

[234] Ibid., 418.

in the immediate aftermath of his confession, utterly disoriented: "'What is this! Where am I!' she said, deeply perplexed, as if she had still not come to her senses. 'But you, you, you're so … how could you make yourself do it?… What is this!'"[235] This prior theory of character also causes Sonya to reject Raskolnikov's initial response to her confused questioning—that he killed simply to rob the pawnbroker: "And how is it, how is it that you could give away your last penny, and yet kill in order to rob!"[236] Thus, rejecting his early explanations, she tries out different hypotheses: "You were hungry! You … it was to help your mother? Yes?"[237] And, later, "The thought flashed through Sonya: 'Can he be mad?' But she abandoned it at once: no, there was something else here."[238] And ultimately, she attributes his behavior to an evil influence: "'Oh, be still, be still!' cried Sonya, clasping her hands. 'You deserted God, and God has stricken you, and given you over to the devil!…'"[239] This idea of the "devil" helps Sonya to reconcile the man she loves with the man who murdered Lizaveta. If he were deceived, if he could be made to understand this, if he would repent, then the Raskolnikov she believes in could be disburdened of his devil and reconnected with God, his family, society, and Sonya herself. Thus, we see that Sonya's theory of Raskolnikov's character generates skepticism about his early explanations for the crime and ultimately leads her to attribute the crime to an external spiritual source.

---

[235] Ibid., 412.

[236] Ibid., 413.

[237] Ibid., 412.

[238] Ibid., 413.

[239] Ibid., 418.

They are face to face for much of the confession, staring into each other's eyes and looking for the truth, as feelings pass contagiously between them. Mental state attributions occur at a much higher rate than verbal communication. And understanding, sometimes achieved in complete silence or in disregard for the words being spoken, is often marked by affect or emotion—such as a bodily trembling that several times marks a realization of some new truth, while preceding any conscious awareness of what has been realized. Sonya is cued to the seriousness of what Raskolnikov wants to say, not by his words but by his mannerisms, and once she understands what he is trying to confess, the incongruence between the morality of the action and the moral character of the actor accelerates this dynamic of mental state attribution and affective contagion. This is most dramatic in the first moment of realization, which is achieved in complete silence. Staring into Raskolnikov's eyes, Sonya finally understands what he has been trying to convey to her—that he is Lizaveta's murderer. The first sign of understanding is affective, as Sonya begins "trembling all over"[240] and unconsciously backs away from Raskolnikov, her arm outstretched defensively, her face fixed in fright. In her bodily expression of fear, Sonya takes on the aspect of the murdered Lizaveta—and then her fear passes into Raskolnikov, initiating a frenzied tango of inferences, affects, and emotions.[241] Raskolnikov's face mirrors Sonya's expression of terror, which likewise mirrors Lizaveta's in the moment before death, "even with almost the same childlike smile."[242] Sonya's whole body responds, without consulting prefrontal cortex. She suddenly "gave a start, cried out, and, not knowing

[240] Ibid., 409-410.

[241] Ibid., 411.

[242] Ibid., 411.

why, threw herself on her knees before him."[243] She is now overcome with pity for Raskolnikov, for the suffering she attributes to him, and this soft-heartedness and this suffering also pass between them. Sonya embraces Raskolnikov and, when he pulls away, cries out that "no one, no one in the whole world, is unhappier than you are now!" before bursting into sobs.[244] In response, "A feeling long unfamiliar to him flooded his soul and softened it all at once. He did not resist: two tears rolled from his eyes and hung on his lashes."[245] As the confession continues, we continue to see how understanding arises through empathy, as they automatically participate in each other's fear, pity, and suffering, all the way to the dramatic conclusion of the scene: "'Sonya, enough! Let me be', he suddenly cried out in convulsive anguish, 'let me be!' He leaned his elbows on his knees and pressed his head with his palms as with a pincers. 'Such suffering!' burst in a painful wail from Sonya'."[246] Out of her overwhelming pain and compassion, in this moment, she tells him what he must do, with words I've already quoted twice.

Thus, we see an exchange of feelings in the confession that proceeds dancelike, in time with mental state attributions, which are oriented towards an understanding of Raskolnikov's character, as the conversation is driven forward by Sonya's assertions that each of his early rationales is "not it." And we see that Sonya's dissatisfaction with these explanations arises from a perceived incongruence between who she knows Raskolnikov to be and what he has done in violating God's least violable law. In short, the confession illustrates all the dynamics of socio-

---

[243] Ibid., 411.

[244] Ibid., 412.

[245] Ibid., 412.

[246] Ibid., 420.

moral cognition outlined in the first part of this chapter. However, these dynamics of rule inference, character attribution, and mental state attribution—facilitated by an exchange of feelings that not only attend but also motivate these cognitive computations, often at an implicit level—are only half the story of socio-moral cognition. As I have argued for much of this chapter, this account of cognitive dynamics doesn't get to the bottom of what all these computations mean and why they are engaged in at all. This account does not tell us all we need to know about the specifically moral part of socio-moral cognition.

Moral meaning is primarily determined by existential framing. Raskolnikov is meaningfully involved with multiple intrinsically good and bad things, in multiple existential frames, and his idea of what is right always shifts in concert with ontological reframings of the people, places, and objects with which he is involved, and correlative reframings of who / what he himself is. Raskolnikov has already constructed most (perhaps all) of these frames privately. They are already there for his perceptions and judgments of things, shifting from situation to situation and mood to mood. What is remarkable in the confession is that we get to witness how these reframings affect Raskolnikov's moral interpretation of his crime, as he tries to explain himself to Sonya. Sonya's disbelief—grounded in her own framing of who Raskolnikov is and what that means for how he ought to behave (character is, after all, an existential frame, though it is far from the *only* relevant frame)—her compassion, conscious and unconscious reactions send Raskolnikov into different existential frames, which give a unique shape to each of his explanations. In his eyes, Sonya takes on the aspect of a holy fool,[247] a judge or priest,[248] an

[247] Raskolnikov has already characterized Sonya and Lizaveta explicitly as holy fools earlier in the novel (p. 324-325), and at various moments in the confession he expresses his continued frustration with Sonya's anti-rationalist appeals to faith. E.g., p. 408: "Once divine Providence gets mixed up in it, there's nothing to be done…" Peavear and Volokhonsky describe a "holy fool" thusly: "a saintly person or ascetic whose saintliness is expressed as 'folly'. Holy

enemy or competitor,[249] a friend or lover,[250] a proxy for Lizaveta,[251] and a child.[252] Raskolnikov himself transforms from a defender of the innocent[253] to a psychopathic criminal[254] to a cowardly "scoundrel"[255] to a desperate son and brother[256] to a visionary great man[257] to a moral

---

fools of this sort were known early in Christian tradition, but in later common usage 'holy fool' also came to mean a crazy person or simpleton" (footnote 12, p. 324).

[248] Ibid., 408: "I was seeking forgiveness, Sonya …"

[249] Ibid., 408: "And suddenly a strange, unexpected feeling of corrosive hatred for Sonya came over his heart." Ibid., 414: "what is it to you if I've now confessed that I did a bad thing? This stupid triumph over me—what is it to you?"

[250] Ibid., 408-409: "he suddenly raised his head and looked at her intently, but he met her anxious and painfully caring eyes fixed upon him; here was love; his hatred vanished like a phantom." Ibid., 412: "'So you won't leave me, Sonya?' he said, looking at her almost with hope."

[251] Ibid., 410: "he looked at her, and suddenly in her face he seemed to see the face of Lizaveta."

[252] Ibid., 410: "she was backing away from him towards the wall, her hand held out, with a completely childlike fright on her face."

[253] Ibid., 408. This is the moral force of the analogy Raskolnikov draws between Sonya's choice to save her family or kill Luzhin and his own choice to save his family from Luzhin by killing and robbing the pawnbroker.

[254] Ibid., 412. I see this in Raskolnikov's explanation that he killed "to rob her, of course." This terse response, with the careless "of course" thrown in, suggests a much more callous and one-dimensional attitude than he expresses in later pages.

[255] Ibid., 414.

[256] Ibid., 415-416.

[257] Ibid., 418. This is the implication of Raskolnikov's claim that he had an idea "that nobody had ever thought before," since such an original vision is the defining feature of his "extraordinary" type of person.

anarchist[258] to a bug,[259] in just twelve pages. And these transformations (along with his reframings of society, the pawnbroker, Luzhin, and so on) determine the essential meanings of the different explanations he gives for his crime: *as* a defender of the innocent and a believer in justice, Raskolnikov is motivated to defend his mother and sister and Sonya (framed as helpless victims) from Luzhin (framed as a vicious predator);[260] as a psychopath, Raskolnikov's motive is simply to steal;[261] as a desperate son and brother and a wasted talent, he seeks to protect his loved ones and himself from the fate imposed by a cruelly impersonal social order;[262] as a spiteful man, he wants to lash out against the stupidity of those around him;[263] as a visionary, he commits the crime to realize his own exceptional nature;[264] as a moral anarchist, he wants to "dare" to break with all morality;[265] and as a disillusioned "louse," he wants to punish himself for his own weakness and lack of originality.[266] Thus, we see that every new rationale for the crime, every change in his account of motive or meaning, is marked by existential reframing of

---

[258] Ibid., 418. "how is it that no man before now has dared or dares yet, while passing by all this absurdity, quite simply to take the whole thing by the tail and whisk it off to the devil! I … I wanted to dare, and I killed … I just wanted to dare, Sonya, that's the whole reason!"

[259] Ibid., 419. "I wanted to prove only one thing to you: that the devil did drag me there then, but afterwards he explained to me that I had no right to go there, because I'm exactly the same louse as all the rest!"

[260] Ibid., 408.

[261] Ibid., 412.

[262] Ibid., 415-416.

[263] Ibid., 417-418.

[264] Ibid., 415, 418.

[265] Ibid., 418-419.

[266] Ibid., 419-420.

who or what he himself is and who or what those involved are, along with reinterpretations of Sonya, society, and so on. Each of these framings has unique moral implications. Each offers a distinct defense or condemnation of his action. Thus, just as essential as all the cognitive abilities and tendencies, which moral psychologists have worn out their keyboards describing—more essential, actually, the sine qua non of moral cognition—is moral existential framing, the determination of moral normativity in view of ever-shifting normative interpretations of beings.

This concludes my analysis of moral polyphony at the proximate analytic level. In view of this characterization of existential framing and moral cognition, I turn in the next chapter to the perspective of ontogeny, considering biological, social, and existential contributions to moral development.

Chapter IV: Moral Development: A Critique of the Idea of Moral Progress

Chapter IV: Contents

History of Developmental Perspectives

The field of moral psychology began as a friendly conversation between philosophy and developmental psychology, in the work of Piaget and, subsequently, Kohlberg. Both men saw consilience between their model of moral psychology and Kant's moral metaphysics. They both observed that the trajectory of moral thought, when given the opportunity and encouragement to develop, was towards individual autonomy and an idea of universal justice. They both observed that the truth that people tended to spontaneously discover as they played and worked together, and still more when they were given time to explicitly reason about moral dilemmas, was that rules are not important in and of themselves but only insofar as they help us relate properly to others. The point of rules is to balance competing interests. The developmental trajectory of moral cognition, it seemed clear, was towards our deliberations becoming grounded in an idea of universal justice. Perhaps through critical dialogue and reasoning we get progressively closer to the moral truth.

A major challenge to this progressivist story came from one of Kohlberg's own students: Carol Gilligan. Kohlberg had developed a system for quantifying moral maturity. Researchers would present participants with moral dilemmas, asking them to make a choice and explain the reasons for their choice and then scoring their answer using Kohlberg's criteria. Gilligan noticed that the women she was studying tended to score lower than the men, and she asked the obvious question: are women generally less morally mature than men? Or is this scale somehow flawed? Gilligan's way of answering this question involved listening attentively to her participants, with an openness to what they had to teach her. What she found was that the women she interviewed sometimes used a different kind of moral logic than the men, a logic that eluded Kohlberg's criteria. She argued that, just because this logic was different didn't mean it was inferior. And

she articulated this different way of doing morality as an equally valid "ethics of care," an approach to moral thought that was oriented towards maintaining relationships and caring for others, as opposed to Kohlberg's approach, which privileged the disinterested application of principles of justice.[267]

Making this simple observation and staking this simple claim was like grabbing a loose thread in the fabric of Kohlberg's theory. Kohlberg had argued that there was a way to gloss over the famous distinction between moral "ought" and factical "is." Tracing how people actually *do* develop under optimal conditions, he claimed, was equivalent to specifying how one *ought* to develop.[268] Gilligan attacked the assumption of moral monism underlying this claim with her suggestion that there might be more than one optimal form of development, more than one optimal set of conditions. In his assumption of monism, Kohlberg was echoing a dominant principle of Western moral philosophy, including the major schools tracing to Aristotle, Kant, and Bentham. In each of these schools, right and wrong are determined in view of a single highest good, which is good universally. The adequacy of moral deliberation can thus be determined with reference to this good. And the acceptance of the proposition of a single highest good means that moral judgments may be hierarchically ranked. Learning how to reason morally means learning how to adequately answer a single question in every situation. But Gilligan claimed that different kinds of people are directed towards different moral questions, different

---

[267] Carol Gilligan, *In a Different Voice: Psychological Theory and Women's Development* (Cambridge: Harvard University Press, 1982).

[268] "These principles [principles of the highest stage of moral reasoning in his theory], I argue, could logically and consistently be held by all people in all societies; they would in fact be universal to all humankind if the conditions for sociomoral development were optimal for all individuals in all cultures" [Lawrence Kohlberg, *The Philosophy of Moral Development: Moral Stages and the Idea of Justice*, Vol. 1. (New York: Harper & Row, 1981), 128].

bases for moral reasoning, and that the different rationalities that arise for these different kinds of people might be equally valid in their own way.

Consider what this means in the perspective of moral development, which Kohlberg and Gilligan, like Piaget before them, conceived as a kind of normative progress. Since Kohlberg thought of moral reasoning as a single competency, he could envision moral progress like a ladder, which people proceed up stepwise, through practice in moral reasoning. But if, as in Gilligan's conception, different people are on different ladders, we must ask whether there is any justification for grouping these different ladders together under a common idea. What is moral cognition anyway? This question arises ineluctably from Gilligan's critique.

A second wave of critique came not long after Gilligan, as Haidt raised this question of the meaning of morality even higher before offering a kind of radical answer. Different cultures, he argued, have different values that build on different moral intuitions. A concern with purity is just different from compassion, which in turn is different from justice. And yet, Haidt pointed out that diverse cultures have developed moral rules pertaining to all of these values and more. It's not just that different kinds of people have their own kinds of morality. Different *cultures* emphasize totally different ways of doing morality. And even within a given culture, you are likely to find several distinct moral values that cannot be synthesized and reduced to a single good. So Haidt expanded Gilligan's criticism of Kohlberg's approach: while Gilligan had attacked the idea that there is a single kind of person and a single best way of reasoning about morality, Haidt expanded this into an all-out war against the notion that there is a single optimal culture that could be maximally conducive to moral reasoning.[269] In fact, there are a diversity of

[269] These assumptions are related, as culture will affect what kind of person one becomes. However, they are distinct insofar as there remains a question of whether there are biological

cultures that each seems convinced of its own ways of reasoning about right and wrong. Kohlberg's whole way of thinking, Haidt pointed out, is contradicted by the facts.

So, with Haidt we have not just two different ladders but several, perhaps *many* ladders, and we find that one's culture strongly influences which developmental ladders one climbs upon and where one ends up as a mature moral subject. We do not, as in Kohlberg's conception, appear to be on a single developmental trajectory, varying only in the quantitative level of moral maturity we reach. And so, the question about morality emerges even more forcefully. What do all these moral ladders have in common that allows us to place their diverse content, all these various values and rules, under the banner of "morality"? What is morality anyway? And how are we to go about defining it?

Here is where Haidt makes a radical move. According to Haidt, we must not look for some uniform moral substance that clearly does not exist, but should instead look at what morality *does*. A "moral system," in Haidt's lingo, is a variety of things that make "cooperative societies possible."[270] Just as there may be different routes to a single destination, there are different ways of creating a cooperative society, different combinations of values and rules, different forms of moral training that are inextricable from the unique way of life of a given social group. Thus, Haidt argued that to understand why we observe a diversity of moral rules and values, we shouldn't try to impose a higher conceptual unity that does not in fact exist. Instead, we should look at how diverse values and principles and practices contribute to some greater outcome for a given group. All moral ladders, in Haidt's conception, lead to cooperation.

(and other environmental) contributions to differences in moral styles, say, between men and women. I'm not going to speculate about this here.

[270] Haidt, *Righteous Mind*, 314.

At least they must have done so within the originary cultural-ecological contexts in which human cooperation evolved. Thus, Haidt argued that to understand morality we needed to pay more attention to human groups, especially on the scale of evolutionary time. Contra Kohlberg, there might be many valid ways of reasoning about morality. All would be justified, as it were, in terms of their conduciveness to cooperation. Thus, Haidt offered a "functionalist" definition of morality, defining moral values in terms of their function of promoting cooperation within a larger "moral system."[271]

Did you notice the basic problem that Haidt's definition fails to address? There are all kinds of things that contribute to cooperation that clearly have nothing to do with morality: like our ability and tendency to imitate others, our intelligence, and various other non-moral competencies. Thus, if we try to define morality in terms of cooperation, we end up including a lot of things that don't belong under the banner of "morality" or "moral cognition." Moreover, on the flip side, there are plenty of things about morality that don't conduce to cooperation, or that conduce to cooperation in some circumstances but work against it in others. Just think of all the bitter divisions within our own society that are caused by moral disagreements. Thus, depending on where / when we look, a functionalist definition of morality might end up excluding a great deal that is clearly part of morality, such as conflicting moral ideologies and values, even as it includes a great deal that is clearly separate from morality, such as intelligence and other competencies. This will not do, and so the question remains how we are to distinguish the moral parts of Haidt's "moral system" from the non-moral parts. If conduciveness to cooperation is a not a definitive criterion for morality, how do we bring together the diverse values that Haidt calls "moral foundations?" What is the rationale for calling them all "moral?"

---

[271] Ibid., 314-315.

This debate between Kohlbergian monism and Haidtian pluralism reveals a basic problem within the field of moral psychology: we don't know how to define morality or moral cognition; still less do we know how to define moral development or moral progress of any kind. The answer to the first question is arrived at by digging down below all moral values to something more basic, as we did in chapter 3, where I showed that morality cannot be defined in terms of values but is first and foremost a response to value that we attribute to beings in the world. Is something "evil?" Resist it. Is something precious? Protect it. Morality is the blanket term describing these kinds of imperatives that emerge out of our attributions of intrinsic value.[272] There may be innumerable forms that such imperatives take, but they all emerge in the same way: through attributions of intrinsic value to things. Thus, my definition allows for a great variety of moral content without violating the thematic unity of the concept of morality.

I accept Haidt's claim that there is no way to reduce the content of diverse moral values to a single thing, such as welfare or justice. But his functionalist definition fails to provide a useful criterion for distinguishing moral parts of a "moral system" from non-moral parts. To find a more adequate concept, we have to dig down below the level of values and systems to the level where morality first emerges. Our sense of responsibility towards valuable beings precedes and underlies any concern with care, fairness, loyalty, authority, and purity (Haidt's "moral foundations"), and much more besides. The latter moral values take shape as our social groups negotiate our moral responses towards valued beings, either by trying to influence our sense of what is valuable or by trying to shift our sense of what constitutes appropriate treatment. So, for instance, we value our parents. Does that mean we should treat them like a close friend, or

[272] We began to see in chapter 3 and will see more clearly in chapter 5 that there are secondary factors that further shape our sense of moral obligation, including the proximity and vulnerability of valued beings. I'm keeping it simple for now.

instead that we should show them a great deal of deference? Different cultures weigh in differently on this issue. But that doesn't mean that one culture is constructing morality, while another is constructing something else. In fact, neither culture is constructing morality from whole cloth. Instead, each culture is participating in a process of cultural *re*shaping and *re*negotiation of the morality of moral beings—individuals who already value parents and many other things, and who already feel that such valuable things ought to be treated in a manner befitting their value.

What does my characterization of morality mean for theories of moral development? Current accounts of moral development are missing the primary thing. Our best theories acknowledge that moral development is an interaction of natural forces of our biology and cultural forces of our society: humans develop morally as we learn the moral forms of our social groups, as we work and play and negotiate with other people. But in focusing exclusively on these two factors, researchers miss out on the bigger picture. It is as if we want to explain how a plant grows, so we focus on the genetics of a plant and a farmer's process of cultivation, through trimming, thinning, weeding, and so on. But then we just leave out the daily processes of sunlight and rain, the rhythms of the seasons and the essential interactions with other organisms. Yes, a plant is born with certain genetic programming. Yes, a cultivated plant is cultivated. But the actual development of a plant, we must appreciate, is an interaction between the organism and *environment* in the broadest sense. Similarly, in moral development, human interactions are only one part of our environmental influence. In reality, morality emerges out of our dynamic relations not only to other people but also to familiar places, sentimental objects, powerful ideas, and beloved creatures.

Is this emergence of morality *progressive*, as in Kohlberg's stage theory? Haidt's critique of Kohlberg, while valid, fails to establish any independent criterion to distinguish possible normative outcomes of moral development (e.g., moral maturity) from non-normative outcomes (e.g., moral immaturity). For this very reason, Haidt cannot refute the idea of moral maturity either. Thus, in this new world of Haidtian pluralism, we have not moved beyond Kohlberg's claims about the possibility of moral progress but have simply abandoned Kohlberg's question about moral maturity altogether. And so, in developing an adequate model of moral development, we must recover this question and ask it in earnest: is there a way to distinguish moral maturity from moral immaturity? I ask and answer this question in the next section.

The Early Development of Morality

Gilligan's and Haidt's critiques of Kohlberg have left us with a big question that the field of moral psychology does not currently know how to address: given the variety of different directions in which morality can develop, is it possible to speak of hierarchical stages of moral development? Is it possible to speak of moral progress at all? On one hand, it would seem not, since development within one cultural system can amount to regression from the perspective of another. Climbing up one moral ladder can amount to climbing down another—as when we develop a sophisticated sense of justice, which causes us to lose the respect we once held for authority. And we lack a universal moral standard according to which we can judge moral maturity and say definitively that "yes, this person has progressed morally," or the converse. In the above example, for instance, a politically liberal person might say we have made moral progress, while a more conservative person might claim we have actually *regressed*. And it is not so easy to provide an objective moral standard that adjudicates between these claims and proves

that, yes, the liberal is right and one has made moral progress, or no, one has not progressed but has actually regressed. However, the fact that our old standards have proven insufficient does not mean that no standard exists. Perhaps we just have to identify the right standard? After all, it would seem strange if virtually every other form of cognition developed progressively, but moral cognition did not.

Consider what in chapter 3 I described as socio-moral cognition. It seems clear that a progressivist, perhaps even stage-theoretical approach is appropriate for characterizing the development of many features of socio-moral cognition, such as theory of mind (ToM), aspects of character attribution, and understanding of moral rules. Tracing the early development of such features, we can witness how they work together, and we can come to understand why some stages must precede others. Looking at early development also helps us to parse out some of the biologically predetermined features of socio-moral cognition from other features that are more culturally or environmentally relative, since the earlier we are in development, the less time our environment has had to influence us. So, by investigating early development, we can get a window into all these features that do develop progressively and into the interaction between biology and environment in shaping our socio-moral cognitive competencies. And we can subsequently frame our big unanswered question about moral development in the context of the existing research in developmental psychology.

Let's begin at the beginning. At very early ages, we see the emergence of pretty much universal capacities and tendencies that are a kind of scaffolding for our mature socio-moral cognitive competencies. Emotional contagion, a precursor of empathy, is one such capacity. Just like many non-human animals, newborn humans of whatever culture come into the world with a tendency to embody the negative emotions of their peers. Put a bunch of two-day-old infants

together, and if one starts crying, the others will form a chorus.[273] Then, from about six weeks, all infants are ready to engage in proto-conversations with willing caretakers, exchanging (typically positive) affects and emotions in successive bids.[274] During the first five years of life, when the balance between the influence of biology and that of cultural learning is tilted in the direction of biology, children develop the ability to sympathize, to take the perspectives of others, to experience self-conscious emotions like embarrassment and shame, to claim ownership over things, to care about fairness in the distribution of goods, to punish and reward others according to their behavior, to understand false beliefs, to dissimulate and lie, and so on. All children in all cultures develop these cognitive capacities and tendencies—though there is increasing cultural variation in the expression of later-developing capacities—and these crucial features of socio-moral cognition develop along a consistent trajectory, one might say, in stages. One has emotional contagion before one expresses consoling behaviors and other, more complex forms of helping that require taking other perspectives, feeling sympathy, and supplying help in a way that is tailored to others' unique needs.[275] One develops an aversion to inequity before one develops more sophisticated ideas of distributive justice, such as ideas of fairness as proportionality, or a willingness to punish non-cooperators at personal cost.[276] One cannot

---

[273] Marvin Simner, "Newborn's Response to the Cry of Another Infant," *Developmental Psychology* 5 (1971): 136-150; discussed by Martin Hoffman, "Developmental Synthesis of Affect and Cognition and its Implications for Altruistic Motivation," *Developmental Psychology* 11 (1975): 614. I note that this response to cries of other infants is significantly greater than infants' response to other aversive stimuli of similar quality and volume.

[274] Rochat, "Layers of Awareness," 138.

[275] de Waal, "Putting the Altruism Back into Altruism," 282-286.

[276] Erin Robbins & Philippe Rochat, "Emerging Signs of Strong Reciprocity in Human Ontogeny," *Frontier in Psychology* 2.353 (2011).

experience emotions like shame and pride until one has developed self-consciousness. And the first explicit claiming of ownership over things also emerges at the same time as children show the first signs of self-consciousness, during the "mirror stage," at around 21-24 months.[277] Similarly, as we saw in chapter 3, one cannot lie before one has developed an appreciation that others can have false beliefs—a landmark ToM competency.[278] Finally, we are only able to use the ToM inferences just mentioned if we have first developed an ability to form useful hypotheses about who and what others are, hypotheses that can help us predict others' behavior and respond appropriately to their needs and desires, but only to the extent that the hypotheses are accurate. Thus, all humans must form useful hypotheses about other's cognitive and dispositional traits from a very early age,[279] and this ability is scaffolding for the subsequent development of ToM capacities.

If all these abilities added up to moral cognition, then a stage-theoretical approach to moral development would be entirely appropriate. There is just one little problem: all these abilities do not add up to moral cognition. These competencies are important in part *because* we are moral beings. But morality is more than this.

As we saw above, moral cognition involves, first and always, subjective value attributions. But it is not easy to rank such attributions in order of normative adequacy. It is in fact impossible. Why? For the same reason it is impossible to rank any subjective judgments.

---

[277] As discussed in Rochat, "Fairness in Distributive Justice," 418.

[278] Talwar & Lee, "Social and Cognitive Correlates of Children's Lying Behavior," 866-881; Ding, et al., "Theory of Mind Training Causes Honest Young Children to Lie," 1812–1821.

[279] These do not need to be subtle or nuanced hypotheses. Even the assumption that others will like what we like is a good starting place. Prior to this, even the presumption that others can have preferences at all—that others have intentions or desires—is extremely useful, and infants show signs of this very early.

Regardless of any consensus we would like to establish as to the "true" value of things, the fact is that people can and do disagree about what is valuable and how valuable valued things are. And we always will because the judgments are subjective. What does it matter if a strict Kantian holds that the only thing that is intrinsically valuable is a rational will? What does it matter if a Benthamite Utilitarian holds that it's rather pleasure? Many otherwise mature adults, perhaps all of us somehow or someway, value intrinsically things other than rationality or pleasure: sacred places, nostalgic objects, beautiful works of art, little babies, or whatever. Is this because we're not morally mature? Who can say? And how could we establish external criteria to measure the "true" value of things, when we are speaking of things whose value is understood as being intrinsic and, as such, not relative to any external standard? The answer is simply that there can never be objective criteria for determining intrinsic value, and thus, there can never be objective criteria for ranking moral judgments. And this line of questioning reveals a paradox at the heart of morality. Even though moral judgment always involves a claim to universality and unconditionality, the fact is that there can be no independent criterion that could adjudicate such claims, simply because there can be no external criterion for determining intrinsic value. And this matters for psychological theories of moral development because it makes it impossible to say in any definitive way that any moral judgment is normatively better than any other and improper to use any such claim to justify a stage-theory of moral cognition.[280] Thus, while we are perfectly justified in proposing a progressivist theory of *social* cognition (along with features

[280] I add a caveat. Since attribution of intentionality is another transcendental condition of moral cognition, and such an attribution partakes in both ToC and ToM computations, it is possible to speak of progress in this *rational aspect* of moral cognition. It's just that there is always an aspect of moral cognition that cannot be said to progress—intrinsic value attribution.

that are necessary but not sufficient for morality, like a ToM capacity[281]), we will never be justified in proposing any straightforwardly progressivist theory of moral cognition. And we should abandon the idea of moral stages once and for all.

Now, just because, as psychologists, we cannot establish a normative stage theory of moral cognition doesn't mean we can't trace how moral cognition develops. We just have to keep in mind what morality is and develop a non-progressivist model of development. Because morality is grounded in subjective value attributions, we cannot say that one's moral sensibility becomes better or more mature, in any objective sense, over the course of development. But one's moral cognition is constantly developing in the sense that it is emerging, fluctuating, taking on complexity and being influenced by experiences, by dialogue with others, by deliberation, by moods, and so on. And we can at least begin to trace this emergence of moral cognition and model how it changes with these factors, starting from a proper understanding of the nature of morality. Moral evaluation emerges out of two forms of normative evaluation: evaluation of things and evaluation of behavior towards valued things. So, tracing the early development of morality means looking at what we know about the development of these distinct forms of evaluation.

Humans are born evaluators, showing differential attraction and repulsion to objects in the world from birth. Even in the womb, fetuses develop preferences for their mother's voice or the smell of the amniotic fluid. We are predisposed to seek comfort and sustenance, with

[281] The reason I have resorted to this awkward expression "socio-moral cognition" is because there are aspects of social cognition that are necessary but not sufficient for moral cognition. The capacity to form a theory of other's minds is one feature of social cognition that is necessary for moral cognition but is not sufficient to constitute moral cognition.

corresponding feelings of warmth and connectedness, which babies evince from early on.[282] Similarly, we don't have to be taught to enjoy sweet foods and (initially, at least) to have an aversion to very sour or bitter things.[283] Through such innately specified preferences, we begin to make distinctions among foods, people, and other things, as we implicitly carve our world into qualitatively specified ontological categories. From the beginning, we understand our world evaluatively, through existential framing.

This means, the groundwork for morality has already been laid by the time we are born. It does not mean, however, that we are born as moral beings. The first step of morality is attributing intrinsic value to things. But just understanding things in qualitative terms needn't involve any attribution of *intrinsic* value, since we might value something only instrumentally, for what we can get out of it. Moreover, just making intrinsic value attributions might not make us moral, since we might conceivably be indifferent towards valuable things. So, to become moral we must first develop a tendency to attribute intrinsic value to things. And we also need to impose upon ourselves and/or others an obligation to be and behave in a way that is appropriate to the intrinsic value we attribute. The first signs of this second-order normativity are the first signs of moral cognition.

I claimed earlier that this rather complex computation, a second-order normativity, must develop quite early, prior to our adoption of moral values and principles, since these latter culturally *re*negotiated forms of morality can only be adopted by beings who already have a subjective moral capacity. Conveniently for me, there is evidence that just such a moral capacity

---

[282] Ann Bigelow & Philippe Rochat, "Two-Month-Old Infants' Sensitivity to Social Contingency in Mother–Infant and Stranger–Infant Interaction," *Infancy* 9.3 (2006): 313-325.

[283] Diana Rosenstein & Harriet Oster, "Differential Facial Responses to Four Basic Tastes in Newborns," *Child development* 59.6 (1988): 1555-1568.

does indeed develop rather early—though, in the absence of a proper understanding of the meaning of morality, this evidence has not been interpreted as radically as it ought to be. Let's consider a 2011 experiment by psychologist Kiley Hamlin and colleagues.[284] This study began with a replication of an earlier finding by the same group, which showed that preverbal infants prefer puppets who demonstrate helpfulness over those who act like jerks (i.e., "hinderers").[285] After replicating the original finding, this new study showed something even more interesting. Splitting infants into two groups, consisting of five-month-olds and eight-month-olds, respectively, the researchers asked whether these babies preferred puppets who helped the helpers or those who helped the jerks; and on the flip side, whether they preferred those who hindered the helpers or those who hindered the jerks. That is, researchers probed for a second-order preference in these infants. What they found was striking, and it seems to me that this study provided evidence of morality in preverbal humans—though Hamlin and colleagues did not themselves make such a radical interpretation of their finding.

Among the five-month-olds in this study, there was no evident second-order preference. These infants preferred the puppet who was helpful, whether it was helpful to another helper or to a jerk. Thus, we can only attribute to these five-month-olds a first-order preference for

---

[284] Kiley Hamlin, Karen Wynn, Paul Bloom, & Neha Mahajan, "How Infants and Toddlers React to Antisocial Others," *PNAS* 108.50 (2011): 19931-19936.

[285] Kiley Hamlin, Karen Wynn, & Paul Bloom, "Social Evaluation by Preverbal Infants," *Nature* 450 (2007): 557–559. In this study, Hamlin and colleagues showed preverbal infants a dramatization involving a puppet trying to climb up a hill. A second puppet subsequently came on the scene and either helped the first puppet up the hill or "hindered" the first puppet by pushing it down the hill. Infants showed a striking preference for "helpers" over "hinderers."

"helpers" and cannot say that this is a moral preference.[286] The eight-month-olds, on the other hand, expressed a second-order preference. They preferred the puppet who helped the helpers, but who *harmed* the jerks. That is, these eight-month-old babies didn't just like those who were nice to others and dislike those who were not; they seemed to like the puppets who treated others appropriately and to dislike those who treated others inappropriately. This latter preference was secondary to these infants' demonstrated preference for helpers over jerks.

Is this second-order preference a moral preference? Under old definitions, it might seem silly to speculate that these babies were expressing morality. For instance, we might try using Kohlberg's definition of morality as "justice reasoning."[287] But using either the term "justice" or the term "reasoning" to describe the preferences of eight-month-olds seems like a stretch. Without being able to ask the infants whether they are applying an idea of justice to the situation, perhaps the best we can do is speculate. Or, we might apply Turiel's definition and seek to determine whether the infants see "hindering" behavior as a moral or merely a conventional violation. To do this, we would need to be able to show that infants view "hindering" as universally wrong, independent of context; however, the above study showed that the preference for helpers over jerks was *precisely* context-dependent for the eight-month-olds, in that it depended on the prior actions of the puppets being helped or hindered. And there is no obvious

---

[286] The specific reason why we cannot say that this is a moral preference is that we cannot say that these 5-mo infants' preference for helping over hindering behavior is derived from a prior value judgment about the things (i.e., puppets) that are being helped or hindered. However, see Hamlin, "Context-Dependent Social Evaluation in 4.5-Month-Old Human Infants," *Frontiers in Psychology* 5 (2014b): 614, for a study that used a more accommodating "habituation" procedure and demonstrated that even 4 ½ month olds can engage in moral judgment under optimal conditions.

[287] Kohlberg, *The Psychology of Moral Development,* 215-217.

way to explain this context by invoking harm, justice, or universal rights, nor any other abstract value, since these participants were so young.

Under my definition of morality, things are different. It seems to me eminently plausible (and I see no compelling alternative) that the infants in this study evaluated puppets based on their actions towards others, such that they initially saw the helpful puppets as good and the unhelpful puppets as bad.[288] Such a value attribution, recall, is the first condition for morality in my formulation. Secondarily, let us suppose that the eight-month-olds developed a normative attitude about how one *ought to* (intentionally) treat puppets, such that one should treat the good ones good and the bad ones bad—without positing this second-order normativity, how do we explain these infants' second-order preferences for those who treated the helpers well and the "hinderers" poorly? If this was indeed the case, in view of the above definition of morality, we can say that these eight-month-olds' preferences were moral preferences, plain and simple. That is, in view of my definition, it seems to me highly plausible that this study provided evidence of moral cognition in eight-month-olds. This should be a big deal. As should the fact that the same researchers recently used a more accommodating "habituation" procedure to help infants understand what was happening in their dramatization, and demonstrated that even 4 ½ month olds perform such judgments, which I am calling moral judgments, under optimal experimental conditions.[289]

---

[288] This assumes the study is well-designed, and the effect is real. I am aware of one critique of their methodology (Damien Scarf et al., "Social Evaluation or Simple Association? Simple Associations May Explain Moral Reasoning in Infants," *PLoS ONE* 7.8 [2012]), but I do not find it compelling. (see Hamlin's response, "The Case for Social Evaluation in Preverbal Infants," *Frontiers in Psychology* 5 [2014a]: 1563).

[289] Hamlin, "Context-Dependent Social Evaluation in 4.5-Month-Old Human Infants," *Frontiers in Psychology* 5 (2014b): 614.

So, the roots of moral cognition—the capacity for existential framing—are already there at birth, and there is now some evidence that perhaps by 4 ½ months, infants already have begun to develop a moral sensibility about how one ought to treat others. From here, there is a lot more to discover about how our sense of the value of things develops and shapes our moral sensibility, along with much that can be discovered about how environmental context influences existential framing and moral cognition. However, my argument above shows that none of this work can support a straightforwardly progressivist stage-theoretical model of moral development. I realize this sounds strange, but the fact is that we cannot say that the moral judgments of a 4 ½ month old are less adequate, less mature, or normatively inferior to those of a 45 year old, though the infant is obviously less mature in other ways. Moral normativity is just not something that can be descriptively said to get better or worse,[290] even as a host of socio-moral competencies do in fact progress. We do grow and progress but, from a descriptive psychological perspective, moral cognition does not, and it is only appropriate to say that this sensibility emerges and changes. What are the factors that influence the emergence and dynamism of moral cognition? This is the topic of the next section, where I engage briefly with sociology.

Environmental Influences on Moral Development

It is a truism that moral cognition emerges through the interaction of biology and culture. But it is not true. This framing blinds us to the reality that many aspects of the environment that shape how humans develop are not social at all. For the sake of analogy, consider physiological development. If we grow up in the Himalayas, we develop large lungs and a tolerance for cold weather. If we get a lot of sun, we get a tan. If we lift weights, we grow muscles. If we sit around

[290] That is, we cannot say this from the descriptive perspective of psychological science. Though naturally we can assert whatever we want.

eating donuts, we turn into a pile of steaming garbage. So, non-social environmental factors like air quality, weather, sun, exercise, and food obviously affect how we develop. The question is whether such factors affect *moral* development. Moral psychologists have always treated morality as a subdomain of sociality, and so moral development has always been understood more narrowly, as an interaction between whatever biological capacity / capacities for morality we are born with and the shaping influence of society. However, if we appreciate that morality is not merely a subdomain of sociality, we can also appreciate that moral development is affected by more than society. Our sense of value is affected by personal experiences with both human and non-human beings. We are born into a social world, but not only a social world, and we develop moral attachments with non-human things too. So moral development should be construed more broadly, as an interaction between biology and environment.

How does this interaction with the broader environment give rise to morality? In chapter 3, I touched on Heidegger's revolutionary reconceptualization of the meaning of being. One of the most important things Heidegger did was simply to point out that an objective, scientific understanding of the world, along with all the things in it, is only one mode of accessing things. And, moreover, that this objective, scientific mode is only ever a secondary way of understanding, after we have already grasped the world in terms of our involvement in meaningful relationships.[291] This is important for moral psychologists to understand, I argued, because, unlike objective properties of things, our existential understanding is full of evaluations, and value is the primary source of morality. So, to understand our moral reality, we have to begin from an existential understanding of things. Heidegger reconceptualizes the categories of experience, such as space and time, similarly, showing that, prior to objectification with

[291] This is a fundamental argument of Heidegger's *Being and Time.*

measuring rods and clocks we already understand space and time in terms of meaningful relations. So, for example, as one is approaching a friend on a road, one is in an existential sense *closer* to the friend than to the path upon which one is walking. And I am closer to these words that I am typing than to the glasses that, for the sake of this example, are perched on my nose. This is a spatial closeness that is defined not objectively but existentially, a spatiality of attention or of "care" in the broadest sense—of Sorge. Heidegger saw that time is like that too. One's past and future are present insofar as one is concerned with them, in all the various way in which we are concerned with our past and future. We do not *exist* discretely in an objective present. Instead, our identity is spread out across the time and space of our meaningful involvements. And even when we are concerned with understanding objective time and space, these objectifications only matter in view of our meaningful activities—working together, going on dates, communing with nature, writing songs, and so on. Thus, prior to objective time, we are already *in* the time of completing projects and participating in activities with others and by ourselves, a kind of time that is existential and meaningful and full of value. This value is the primary source of morality.

In making this claim, I am critiquing the dominant view in moral psychology, which attributes moral development to an interaction of biology and socialization. But my view can be read as an expansion of the accepted story, rather than a simple refutation. Indeed, there are some striking similarities between the existential thinkers I draw upon—Dostoevsky and Heidegger—and the great sociological theorists of morality. For instance, it is remarkable that, prior to Heidegger, the sociologist Emile Durkheim had already developed something closely analogous to Heidegger's idea of existential worldhood, spatiality, and temporality. Like Heidegger, Durkheim rejected the notion that time and space are primarily objective, and he pointed out that

prior to objectification, we have already given things a human meaning, a meaning that expresses our own goals and aspirations, our own ways of life, our own desires and needs as human beings.

The major difference between this interpretation and that of Heidegger is that, in his obsession with sociology, Durkheim interpreted all things in terms of human society. The universe itself was, for Durkheim, "part of society's interior life."[292] And the categories of experience, such as space and time, were defined by society and essentially corresponded to the places and rhythms of social interaction.[293] Heidegger is more expansive and, obviously, I consider this expansiveness more adequate than Durkheim's social reductionism. It is not society that gives us our existential spatiality and temporality. Rather, it is in our very nature to understand time and space in terms of our meaningful activities. These activities may be social, but they may also be non-social or extra-social. Or there may be dimensions of our meaningful activity that are not social, existing alongside the social dimension. We cognize in both conceptual and pre-conceptual modes. We are meaningfully involved with people, places, creatures, objects, and ideas, among other things. Thus, whereas for Durkheim the concept of the universe, or indeed of any totality, is "only the abstract form of the concept of human society,"[294] for Heidegger, the concept of "world" is inclusive of all things with which we find ourselves meaningfully involved, all things that show themselves to us as things. Or simply, all things. Society with humans might be a privileged form of existence, a mode of being that strongly influences most of our encounters, even with non-human beings. But this reality should not blind

---

[292] Emile Durkheim, *Emile Durkheim on Morality and Society: Selected Writings*, ed. Robert Bellah (Chicago: The University of Chicago Press, 1973), 217.

[293] Durkheim, *Morality and Society*, 217-218.

[294] Ibid., 217.

us to the many ways in which non-human beings assert their reality in our world—a reality that is not reducible to human society, to language, or even to conceptual ways of understanding.

How does Heidegger's more expansive ontology support a more adequate characterization of everyday morality? Consider the stars. Whereas for Durkheim, external changes, such as seasons or the changing positions of stars are meaningful only insofar as they make an "essentially social organization intelligible to all,"[295] for Heidegger, it is important to allow such things to show themselves in their own way. Far from being reducible to some signal of time that humans rely upon to coordinate their activity with other humans, seasons impose their own intensely physical and never perfectly predictable realities upon human existence. Stars may help even isolated wanderers to find their bearings. Or they may speak of a reality that truly is much greater than human life: not because the stars are a Durkheimian image of the fact that society is greater than the sum of its human parts but because, for instance, the stars may speak of the insignificance of *all* human existence, the insignificance of the entirety of human history. Such a claim might be made by stars when we take them for more than mere images of ourselves, when we allow them to speak in their own voices. If we listen, we will hear the melodies of the stars within our own moral polyphony.

In privileging the social dimension of the environment, moral psychologists have on the whole failed to listen to the voices of the stars and other non-human beings. This is where an existential analysis is helpful. Let us take this claim about the stars seriously and ask why: Why did the Catholic Church ban Copernicus's writings for hundreds of years and threaten Galileo with torture over his heliocentrism? Or, forget the stars. Why is Darwinian evolution still resisted today by the religious right? The answer is quite simply that the insights of Galileo and

[295] Ibid., 218.

Copernicus and Darwin challenge a moral order that places humans at the center of the universe, and that sets humans ontologically above other life forms. Of course, we know this. It is obvious. But moral psychologists who treat morality as something that emerges through an interaction of biology and culture have not appreciated what this says about morality. Our sense of morality is affected by our personal encounters with animals and plants and stars and landscapes and ideas. Morality is affected by insights in domains of physics and evolutionary theory. Morality is not like some program that we download from our cultural server, or some cultural competency that we develop simply through interactions with those around us. Rather, our moral sensibility is constantly emerging through our encounters with humans and non-humans, as we form attachments, as we make discoveries, as we create meaning, as we tell stories, both socially and privately, about who we are in relation to all things. Just by observing the world, just by engaging in meaningful activities, forming relationships with people, places, creatures, objects, and ideas, we are constantly forming and reforming our sense of what is valuable and how we ought to treat valued things. We are constantly forming and reforming our moral sensibility. So, while it is true that through dialogue with other people this moral sense get renegotiated and often, as it were, *leveled out* to fit more or less comfortably within our cultural milieu, it is also true that we can only learn a particular cultural moral dialect because we are *already* moral beings. And even as we negotiate morality with others, we never stop developing unique personal attachments that exert their own moral force. My existential analysis reclaims the reality that moral cognition is unique to each individual and, indeed, every moment, a reality that is inaccessible to theories that acknowledge only the interaction of biology and culture.

Only against the backdrop of this broader understanding of how our moral sensibility emerges through personal interactions with beings in the world can we properly trace the role of

society in shaping moral development. In chapter five, I am going to elaborate on how our moral sensibility emerges through interactions with a broader environment, incorporating the perspective of "situationism" in my model. Situationism is typically framed as a negative theory, undermining philosophers' assumptions about moral responsibility by showing that humans have much less agency than we tend to assume. But situationism can also be incorporated into a positive theory that treats our relations to the broader environment as constitutive of the moral domain itself: instead of pointing to an absence of agency, the situationist literature shows how agency actually emerges from moment to moment, through changes in existential framing. Thus, I will incorporate the situationist literature into my polyphonic model of moral psychology.

Now I am almost at the end of this chapter on moral development, in this Dostoevskian model of moral psychology, and I have not invoked Dostoevsky at all. This was not my original plan. I was going to read a passage from Dostoevsky's novel *Demons*, where he develops his critique of Eurocentric ideas of moral progress. But I decided that, while Dostoevsky's understanding is consistent with the Heideggerian account I developed above, it is easier for me to develop this model without engaging too much with Dostoevsky's work. Thus, at this point, I will simply offer a few examples in support of this claim that Dostoevsky's model is consistent with the above account of development as a non-progressive interaction between individuals and environment in a broad sense, and not merely a progressive interaction between biology and culture.

For Dostoevsky, God is not just an idea that society forms of itself, though he does express something very close to this proto-Durkheimian idea in several places.[296] When we look

[296] Within Dostoevsky's fiction, this idea is most elaborated in Nikolai Stavrogin's last conversation with Ivan Shatov in *Demons* (Fyodor Dostoevsky, *Demons*, trans. Richard Pevear & Larissa Volokhonsky (New York: Alfred A. Knopf, Inc., 2005), 250—256.

at Dostoevsky's work in all its details, it is clear that the idea of "God" can express one's relation to the whole world, stars included.[297] And moral cognition is influenced by all kinds of environmental or situational factors, beyond simply our relations to other people. For instance, Raskolnikov tells Sonya that his crime owed something to a state of mind that was induced by his physical surroundings: "low ceilings and cramped rooms cramp the soul and mind!"[298] Thus, Raskolnikov suggests that physical space shapes moral cognition. In Dostoevsky, we also see that idleness tends to correlate with particular kinds of morally relevant pathologies, as it does for Raskolnikov, the Underground Man, Stavrogin, and Fyodor Karamazov; whereas industry is correlated with virtue and faith in characters like Dmitri Razumikhin from *Crime and Punishment* and Alyosha Karamazov and the elder Zosima from *The Brothers Karamazov*. Finally, we see a connection between the physical ailment of epilepsy and mystical spiritual beliefs in prince Myshkin and in the existential suicide from *Demons,* Alexei Kirillov. Consider all these factors. Idleness, industry, physiological illness, the size of one's living quarters—these are not social facts. Even when such factors are affected by social realities, they remain essentially non-social situational features. And they are part of Dostoevsky's story of moral development and change. Thus, we must acknowledge that society is not the whole story for Dostoevsky, and that he also appreciates a broader existential sense of the meaning of moral life, which is consistent with the Heideggerian account I have been developing: an account that understands moral development in a non-progressivist sense as an interaction between individuals and environments in the broadest sense.

---

[297] This more expansive view of God is perhaps most eloquently expressed by Father Zosima, in *The Brothers Karamazov*, but it is also expressed by Prince Myshkin and Sonya Marmeladov, among others.

[298] Dostoevsky, *Crime,* 417.

This is all I'm going to say about Dostoevsky in this chapter. His work is consistent with my model of moral development. In chapter five, I will do a final reading of *Crime and Punishment*, tracing the role of the situation in Raskolnikov's crime and confession.

Chapter V: Morality & The Situation

Chapter V: Contents

"This moment, as it felt to him, was terribly like the one when he had stood behind the old woman, having already freed the axe from its loop, and realized that 'there was not another moment to lose'. 'What's the matter?' Sonya asked, becoming terribly timid. He could not utter a word. This was not the way, this was not at all the way he had intended to announce it, and he himself did not understand what was happening with him now."[299]

The Parable of the Daisies

Imagine there is a gardener whose job is to water the flowers of two large estates, one very near her home and the other far away. The gardener always waters the flowers at the nearby estate in the morning and makes her way out to the distant estate to water those flowers in the evening. Noticing that the daisies in the nearby estate always face east and those in the distant estate always face west, our gardener concludes that daisies are dimorphic, having one of two distinct traits: one that causes them to incline to the east and another that causes them to incline to the west.

Our gardener tells her friend, a botanist, about this dimorphism of the daisies. But the botanist is skeptical and offers a different explanation. Maybe there is no internal trait that causes daisies to tilt east or west. In reality, the direction of their tilt might depend upon some external factor that varies between the two situations in which the gardener happens to encounter the daisies. Maybe the tilt of the flowers is influenced by the topography of the land where they are planted or the direction that the wind tends to travel on each estate at the time when the gardener sees them. Maybe this behavior has nothing to do with the daisies' internal traits but is really caused by some external feature(s) of the situation.

What is wrong with this debate? Both the gardener and the botanist are somehow ignorant of the fact that daisies are heliotropic, meaning that they tilt their faces to follow the sun. Thus, in forming their alternative hypotheses, the gardner and the botanist are both partly

[299] Dostoevsky, *Crime*, 109.

right but basically wrong. Obviously, their specific explanations are wrong. The gardener's hypothesis is inconsistent with the reality that daisies on both estates face east in the morning and west in the evening. And the botanist's hypothesis is inconsistent with the fact that the sun's light elicits this response only from flowers that possess the trait of heliotropism. Despite being wrong in the specifics, each of these hypotheses touches on part of the truth: consistent with the gardener's hypothesis, the daisies' behavior really is determined by an internal trait (heliotropism); consistent with the botanist's, the behavior really is determined by an external environmental feature (the sun's movement). The basic error of the debate is the assumption that causation must be *either* internal or external, and the attendant failure to consider a third possibility: that an internal trait of the organism and an external feature of the environment might work together to produce the behavior.

Moral psychologists have been engaging in a debate that in some ways resembles that between the gardener and the botanist. We have been arguing for decades now over what's more important for moral judgment and action: internal traits of the person or external features of the environment. While some interesting findings have animated this debate, we have failed to properly interpret them. The opposition between external "situational" features and internal character traits is mostly superficial. At a deeper level, the external and internal work together to produce moral / immoral behavior. However, we have missed the significance of this simple truth simply because we have not identified this deeper trait—the originary source of moral behavior.

What is this mechanism that has been hiding all these years? We saw in chapter 3 that the originary source of moral behavior is our responsiveness to changing qualities of *being,* as our ontological understanding shapes our moral sensibility. When we identify a person, place,

creature, object, or idea as being intrinsically valuable, we tend to feel a sense of moral responsibility to respect it or protect it—broadly, to treat it properly. And the extent of our moral responsibilities is further shaped by the personal relevance of the thing: how existentially close it is to us, how much a part of our selves and our world it is. Finally, we must take greater care with more vulnerable things. Thus, there are basically three aspects of being that shape our sense of right and wrong behavior: value, proximity, and vulnerability.[300] Just as daisies are responsive to the intensity and the angle of sunlight, moral beings are responsive to the intrinsic value, the personal relevance, and the vulnerability of beings.

Let us consider each quality in turn. Intrinsic value is the most basic condition, the sine qua non of existential framing, morality, and moral cognition. Just as the sun must shine in order for heliotropic behavior to emerge at all, so must we attribute intrinsic value to beings in order for moral cognition and behavior to emerge at all. The less valuable we perceive something to be, whether because we see it as ugly or false or cheap or whatever, the less compelled we will be to respect or protect it. Thus, in the same way that a shadow falling over the daisies might attenuate the light enough to reduce or eliminate their expression of heliotropism, our devaluation of beings can reduce or eliminate our sense of moral responsibility towards them. Next, consider personal significance. Just as the angle of the sun's light orients the daisies, our sense of closeness, affiliation, or personal interest orients our moral sensibility. Thus, personal relevance is a cue like the angle of the sun's light. We feel a strong sense of moral obligation towards things that are near to us, in the existential sense of nearness, and we feel less responsibility towards things that are distant, things that are, in any number of ways, "not our problem." We hardly feel the moral pull of beings that are sufficiently far away, beings that we

[300] Perhaps there are others that I haven't thought of?

have actively or passively excluded from our circle of affiliation and identity. And just as the daisies' faces close at night, so our moral concern is closed off from beings when we ignore or forget them. Vulnerability is a bit more complex. It's not that we necessarily feel a greater moral obligation towards vulnerable beings but, rather, that the *quality* of our obligations may change. We might feel obligated to *protect* vulnerable things like babies and, conversely, to *obey* powerful entities like gods or monarchs.

In view of the nuance involved in a discussion of vulnerability, I am going to focus on value and proximity in the following discussion. I summarize this claim about the moral significance of value and proximity by saying that our sense of moral responsibility is elicited by what is near and dear to us. This responsiveness to dearness and nearness is an internal trait that we have as moral beings: I have described it as a capacity for existential framing. I grant the situationists' point that any assumption that moral cognition happens in a vacuum, that people reason about right and wrong and are not moved by the situation, is clearly contradicted by empirical evidence. However, in arguing that the external environment determines much of our behavior, "situationists" have not acknowledged the mediating role of this internal capacity for existential framing upon moral cognition and behavior. Thus, I claim that both the situationists and their opponents are partly right but basically wrong. I do believe that external situational factors determine much of our behavior. But I show that these factors do so primarily by engaging an internal trait: our capacity for existential framing. The most interesting and impressive experiments cited by "situationists" involve manipulations of the dearness and/or nearness, the value and/or personal relevance of their independent variable(s). What such experiments show is thus the power of existential framing, the malleability of our sense of right and wrong in response to our changing understanding of the value and proximity of things. A

proper interpretation of these experiments thus synthesizes the opposition between internal and external, showing how the influence of the situation on moral cognition is largely mediated by the internal capacity for existential framing. Below, I offer just such a critical reinterpretation of the situationist literature.

Situationism

As alluded to above, situationism is a psychological model of human behavior that emphasizes the importance of environmental factors over personal qualities for determining behavior. In moral psychology, this view has been used to critique an assumption of moral philosophers and everyday folk: that people act out of a free will, in accordance with their unique character and desires.[301] Citing evidence of a strong influence of environmental factors upon moral behavior in numerous psychological studies, situationists argue that our actions mostly do not reflect stable character traits. We are not rational and autonomous and unique, as many philosophers presume or demand, but are easily swayed by accidental features of our environment. Our actions are determined, to some surprising degree, by our external situation. This is what I will call the "classic" situationist interpretation of moral psychology.

Some of the studies in the situationists' cannon show that moral behavior can be modulated by the introduction of an unpleasant smell or sound, by giving participants an unexpected windfall, and so on.[302] We might call these "orthogonal" situational factors, since they change one's moral behavior even though they seem to have no direct relevance to morality.

---

[301] See John Doris, *Lack of Character: Personality and Moral Behavior* (New York: Cambridge University Press), 2002.

[302] Appaiah, *Experiments in Ethics*, 40-42.

The classic situationist interpretation, paraphrased above, is basically designed to explain the influence of these factors. Just as wind might bend the daisies without engaging any internal trait, so orthogonal situational factors might influence our behavior without engaging our internal moral-cognitive traits.[303] If all the studies cited by situationists demonstrated a strong influence of orthogonal situational factors, the classic situationist interpretation would be pretty good. I would basically buy the argument that external factors tend to be more decisive for moral behavior than stable character traits and rational processes—while acknowledging that the latter are also important. However, the reality is not so simple. The most interesting, impressive, and (in)famous studies in the situationists' cannon generate dramatic behavioral effects by changing participants' sense of the nearness and dearness of other beings. Situational factors that work in this way are not orthogonal to morality at all: they change behavior by reframing one's understanding and thus actually changing one's sense of right and wrong. In these latter experiments, behavior modification is mediated by a core internal feature of our moral psychology: the capacity for existential framing.

My first exhibition of this latter dynamic is the infamous Stanford Prison Experiment of 1971, wherein participants were randomly assigned to be either guards or prisoners in an intensely realistic, weeks-long role-play. Under these conditions, psychologist Phillip Zimbardo showed that regular people (the "guards") could be induced to exhibit shocking aggression

[303] I add the caveat that sometimes factors that appear orthogonal might actually engage our internal sensors of nearness or dearness: if we are feeling grateful, because we have just found some unexpected cash, a bump in our mood could cause us to look at others more favorably and thus to be more generous out of a boost to our dearness sensor; if we are in a hurry to get somewhere, we might not take the time to affiliate with a stranger in need, and so factors that induce hurriedness might cast shade over our proximity sensor.

towards innocent strangers (the "prisoners").[304] This experiment is often cited in support of the classic situationist interpretation of behavior. It shows that what many of us regard as immoral behavior might arise not because of internal differences in people's moral character—after all, guards and prisoners were assigned randomly to their respective groups—but, rather, because of external differences in the situation. However, in making such an interpretation, one ignores the possibility that the influence of external situational features might be mediated by an internal trait (a trait that happens to be shared across participants); and we thereby make an analogous error to that of our ignorant botanist, who did not consider the possibility that the daisies' tilt might be mediated by an internal trait such as heliotropism.

If the relevant situational factors in Zimbardo's experiment are not "orthogonal," what are they? The roles of guardhood and prisonerhood are existential frames, and one's assignment to one role or another affects one's behavior by engaging one's *internal* capacity for existential framing. In this experiment, we can easily identify which aspects of being are modulated by these assignments: Zimbardo manipulates *both* proximity and value, both nearness and dearness. As a guard, one feels a sense of affiliation with other guards and a sense of shared goals. There is a job to be done, and it involves working together with other guards to manage the prisoners. From the guards' perspective, prisoners are the out-group. They are cordoned off physically and categorially. They do not have the same goals or interests as the guards—quite the opposite. So, the sense of existential nearness or shared identity is attenuated across the two groups, even as it is strengthened within groups. Beyond this manipulation of existential proximity, there is also a manipulation of value. Prisoners are a devalued group. A prisoner is someone who has been stripped of some of their rights, typically as punishment for illegal (and perhaps "immoral")

[304] Craig Haney, Curtis Banks, & Philip Zimbardo, "Interpersonal Dynamics in a Simulated Prison," *International Journal of Criminology and Penology* 1 (1973): 69-97.

behavior. From the perspective of a guard, a prisoner might be seen as a "bad guy," deserving of what they get. So, the value of "prisoners" is likely also attenuated, or even inverted from good to bad, in the minds of guards.

The behavioral effects Zimbardo reported are consistent with either or both of these ontological transformations, these dramatic existential reframings of the people in these two groups. Zimbardo's elaborate production induced realistic changes in both the nearness and dearness of one group relative to another, engaging participants' *internal* capacity for existential framing to achieve dramatic behavioral effects. Thus, it really makes no sense to conclude that these effects were the result of orthogonal situational factors bending the behavior of participants as the wind bends the daisies. Just as the sun tilts the daisies by engaging their internal attribute of heliotropism, Zimbardo achieved his effects by engaging with an internal attribute of all moral subjects: the capacity for existential framing.

A similarly impressive example of situational influence on behavior is supplied by Stanley Milgram's so-called obedience studies. Back in the 1960's, Milgram showed that everyday people are willing to shock innocent strangers at what they believe to be excruciating and dangerous voltage when instructed to do so by an authority figure. As with Zimbardo's experiment, Milgram's studies have generally been interpreted as supporting the classic situationist interpretation of moral behavior. In the wake of WWII, as the western world was trying to come to grips with the Nazis' crimes, Milgram's findings suggested that a majority of everyday Americans might be just as willing as Nazis to commit atrocities out of blind obedience to authority. Perhaps at the level of internal character traits, Americans and Nazis are equally malleable, and it is only the better external conditions of our society that stop us from falling into similar forms of depravity.

While these experiments make a compelling case about the potential for everyday people to commit violence, Milgram misrepresented the mechanism through which this occurs. A closer look at his studies reveals that participants were not at all blindly or passively obedient to authority. For instance, participants in Milgram's experiments, and in more recent replications, were universally resistant to the prompt experimenters sometimes used as a last attempt to induce cooperation: "You have no other choice. You must go on."[305] No one ever obeyed this command, but chose instead to rebel in the face of explicit coercion.

If not blind obedience, what induced participants to engage in such shocking behavior? Psychologists Stephen Reicher and Alexander Haslam argue that, rather than being passively bent by the wind of authority, participants responded to numerous cues that shifted their proximity, their social alliances in the context of the experiment.[306] For instance, compliance with the direction of the experimenter was brought down by all factors that took subjects out of the frame of shared participation in an important scientific project, whether by emphasizing their subjection to the authority of the experimenter or by, in various ways, delegitimizing the scientificity of the experimental process.[307] Factors that helped place participants in a common frame with the shock victim, such as increasing their physical proximity, also brought down compliance.[308] Conversely, compliance was increased by factors that emphasized subjects'

---

[305] Stephen Reicher & Alexander Haslam, "After Shock? Towards a Social Identity Explanation of the Milgram 'Obedience' Studies," 2011, 167-168.

[306] Of course, I would replace the word "social" with the more expansive term "existential."

[307] Reicher & Haslam, "After Shock."

[308] Ibid.

participation, together with the experimenter, in an intrinsically valuable scientific project.[309] Thus, it is clear that Milgram achieved his most dramatic effects through three distinct manipulations of existential framing: 1. attenuating the sense of proximity between participants and the person they believed they were shocking, 2. strengthening participants' identification with the scientific project, and 3. buttressing participants' sense of the value of the research they were supporting. These behavioral effects were thus mediated by participants' internal capacity for existential framing, as they responded to numerous experimental manipulations of proximity and value. Thus, in the situation in which they found themselves, participants were not simply going against their conscience out of a lack of character strength or agency. Rather, they found themselves in a situation where their obligations as participants in an important scientific experiment conflicted with their obligations towards another study participant. Compliance with the direction to shock the other participant was not an abnegation of moral responsibility but a particular moral choice, a performance of responsibility in a morally complex situation.

With this brief discussion, I want to suggest that existential framing is of primary importance for determining how the situation affects moral cognition and behavior. I have a couple reasons for feeling that orthogonal situational features are less influential than situational features whose influence is mediated by existential framing. For one, Milgram's and Zimbardo's experiments are far more impressive than most of the experiments that rely on orthogonal situational factors. These latter studies typically show small or medium effects upon low-stakes giving behavior—big deal![310] Moreover, as I mentioned in a footnote above, some of the influence of even "orthogonal" manipulations may be attributed to their indirect influence on

---

[309] Ibid.

[310] Jk. There are some cool findings.

existential framing: for instance, factors that manipulate our attention might interrupt our affiliation or concern and thus preemptively interfere with existential framing, like a cloud that interrupts the daisies' movement by temporarily blocking the sunlight. Thus, while I acknowledge that some purely orthogonal factors influence moral cognition and behavior in a way that is consistent with the classic situationist hypothesis, I propose that these cases are of secondary importance. Situational change influences moral cognition and behavior primarily by influencing our understanding of the nearness and dearness of beings. To show this dynamic in action, I now turn to a final reading of *Crime and Punishment*.

Raskolnikov and the Situation

Raskolnikov is a divided person, full of contradictions, and I have been arguing that, in fact, we are all like that, even if we don't realize it. We are polyphonic. Moral polyphony grows from the very roots of morality, from how our ontological understanding of things gives rise to our sense of moral responsibility. We have moral relationships with many things: these are the voices within our moral polyphony, each valued being calling out for our consideration. All valued beings make such demands, and giving proper consideration to one being often involves giving less consideration to others. Even our sense of what constitutes proper consideration changes as we move through the world, as we form or break attachments, as our attention shifts from one thing to another, even as our moods change. The voices compete. Voices may be drowned out by noise in the environment, or by the clamor of other voices. Voices may grow more distant or more insistent. And voices speak to us even when we do not realize we hear them, like this melody that's been playing in my head for days.

As we saw above, the situation affects moral cognition primarily by modulating these voices. For instance, if I am induced to pay more attention to something because it is brought closer to me or made more salient, its voice will be turned up in the mix of my moral consciousness. If I forget about something, its voice will be turned down in the mix, and I'll also forget about my moral responsibilities towards it. When this process happens implicitly and automatically, moral cognition and action will tend to be parochial, as we give moral consideration to what is right in front of us, while simply ignoring whatever is not so immediate. But parochialism can also be conscious and deliberate, as we see so strikingly in nativist political movements around the world today. At times, we may resist such parochialism by consciously listening to the faint voices, by bringing the distant things into the circle of our moral consideration. We may try to maintain moral consistency, acting according to principles that remain the same across different situations. But the situation may undermine our attempts at consistency, especially when it affects us at a subconscious level.

Whatever the conscious and unconscious, internal and external influences upon moral cognition, and whatever the outcomes of our moral thought, whether we believe in universal rights or espouse a more parochial form of morality, we should bear in mind that the same fundamental process of moral cognition is engaged in every case. Whether moral cognition is conscious or unconscious, whether it involves free agency or not, whether it is principled or unprincipled, it always involves this correlated emergence of ontological understanding, value attribution, and felt moral responsibility. Just as our breathing may change from moment to moment but always involves the same underlying process of respiration, so moral cognition may change from moment to moment but always involves the same underlying process of existential framing. We can take conscious control of our moral cognition, as when we consciously control

our breathing, but most of the time we just do morality automatically. And all kinds of situational factors affect how we do it. Just as breathing is automatically affected by whether we are lying down or running up a hill, whether we are in shape or not, whether we are at high or low altitude, and so on, moral cognition is automatically affected by our proximity to valued things, our habits and training, our changing mood, our locus of attention, and so on.

To talk about Raskolnikov's changing existential framing of himself, Sonya, the pawnbroker, Luzhin, and Russian society is also implicitly to talk about the situation. However, these reframings might appear to be entirely motivated from within, with Raskolnikov creating rationales for his crime, as it were, from whole cloth. In my earlier readings of the confession, I traced these changing existential frames without looking closely at the relation between Raskolnikov's internal processes and the influences of the external situation. This could mislead one into thinking that Dostoevsky's idea of morality is a kind of solipsistic existentialism, where moral subjects simply invent moral responsibility for themselves. While Dostoevsky does argue strongly for freedom, agency, and personal responsibility, his argument does not involve any denial of the influence of situational features. And the normative argument in favor of moral responsibility, which Dostoevsky makes implicitly in *Crime* and explicitly elsewhere,[311] is always carefully separated from his nuanced descriptions of how things really are. Dostoevsky recognizes that the value of taking moral responsibility is totally independent from the answer to the question whether we are *ultimately* responsible for our actions or not. Thus, despite his prescriptions that we ought to take responsibility for our actions and, indeed, even for the actions

[311] In his *Writer's Diary,* for instance, Dostoevsky railed against what he saw as a dangerous trend of Russian juries excusing criminals due to mitigating factors of the "environment." Fyodor Dostoevsky, *A Writer's Diary: Volume One,* trans. Kenneth Lantz (Evanston: Northwestern University Press, 1993), 1873, 3, "Environment," 132-145.

of others,[312] when it comes to describing behavior, Dostoevsky is very attentive to the influence of the external situation upon and even over the individual. Skeptical? I'll prove it, rounding out my account in this chapter by tracing the many ways in which Raskolnikov's moral thought and action are shaped by situational features of the environment.

As I've maintained since chapter 1, Dostoevsky's portrayal of how thoughts happen is consistent with our contemporary understanding in cognitive psychology that much of thought happens implicitly, outside of our conscious control. His descriptions are also consistent with the view that situational factors strongly influence both conscious and unconscious (but especially unconscious) cognitive processes. We see this very clearly in Raskolnikov's sudden inspirations, periods of dissociation and forgetfulness, moments of uncontrollable rage, and so on, which are typically connected to external environmental influences. Dostoevsky appreciates that even internal processes are often outside one's control. Thoughts, in Dostoevsky's portrayal, are not always something we, as agents, do. Thoughts often *happen to* us, and the way in which they happen is strongly influenced by the situation. Thus, after Raskolnikov has read that distressing letter from his mother and realized that in his present circumstances he is powerless to stop his sister's marriage to the pig Luzhin, suddenly the thought he had been thinking already—about killing the pawnbroker—comes back unexpectedly and "hit him in the head."[313] The thought that had been a mere dream is transformed by the situation: "now it suddenly appeared not as a dream, but in some new, menacing, and quite unfamiliar form."[314]

---

[312] Dostoevsky, *Writers Diary: Volume One*, "Environment."

[313] Ibid., 45.

[314] Ibid., 45.

In this case, the situation drags Raskolnikov towards the crime, but there are really two directions in which Raskolnikov is influenced, and so we might say there are two situational forces: a "devil" dragging him towards the crime and an "angel" pulling him in the opposite direction. At moments, the angel seems to succeed in liberating Raskolnikov from his intention to kill. But each time the devil brings some other situational coincidence to bear, dragging Raskolnikov back onto the path. The most dramatic and fateful of these vicissitudes occurs the day before the crime. After falling exhausted and half-drunk into some bushes and dreaming vividly of an incident from his childhood, when he had witnessed a drunken peasant beating his horse to death, Raskolnikov awakes with a feeling of strong revulsion against the idea of committing violence. This revulsion, evidently brought about by his unconscious mind under the influence of sleep deprivation, illness, and alcohol, seems to liberate Raskolnikov from his murderous intention:

> He got to his feet, looked around as if wondering how he had ended up there, and walked towards the T—— v Bridge. He was pale, his eyes were burning, all his limbs felt exhausted, but he suddenly seemed to breathe more easily. He felt he had just thrown off the horrible burden that had been weighing him down for so long, and his soul suddenly became light and peaceful. "Lord!" he pleaded, "show me my way; I renounce this cursed … dream of mine!"[315]

The devil does not let Raskolnikov go that easily, however. Lost in thought, walking without charting the course, he unconsciously takes a circuitous route home, passing through the Haymarket. Here, altogether by chance, Raskolnikov overhears the pawnbroker's sister Lizaveta speaking to some tradesfolk and agreeing to meet them the next day between six and seven in the evening. Now that the devil has shown him the precise time when the pawnbroker will be completely alone, Raskolnikov suddenly no longer feels free *not* to go through with the plan.

---

[315] Ibid., 59-60.

This little circumstance drags him forward: "He walked like a man condemned to death. He was not reasoning about anything, and was totally unable to reason; but he suddenly felt with his whole being that he no longer had any freedom either of mind or of will, and that everything had been suddenly and finally decided."[316]

Now this idea that had been little more than a dream takes on even more reality and urgency. The letter from Raskolnikov's mother telling of the situation with his sister, his feeling of utter helplessness to do anything to stop this marriage from happening, his physical and psychological deterioration, and this sudden opportunity that seems given by fate: these circumstances conspire in dragging him towards the crime.

This sense of being dragged by the situation recurs at every crucial stage of the crime and the cover-up, from the beginning. For instance, after his first meeting with the pawnbroker, as he is beginning to conceive the idea of the murder, Raskolnikov overhears a conversation about this very same pawnbroker in which one man makes a moral argument in favor of killing the old woman, an argument that echoes Raskolnikov's own thoughts at that very moment. This circumstance seems fateful to him:

> Why precisely now did he have to hear precisely such talk and thinking, when … exactly the same thoughts had just been conceived in his own head? And why precisely now, as he was coming from the old woman's bearing the germ of his thought, should he chance upon a conversation about the same old woman?[317]

Raskolnikov does not yet feel that he is being dragged by the devil. But this is the first shadow of the situation, which comes by the end to impose itself so powerfully.

The situation exerts influence before, during, and even after the crime. Dostoevsky reveals this devil of a situation little by little, in all sorts of details. For instance, it so happens

---

[316] Dostoevsky, *Crime*, 62.

[317] Ibid., 66.

that the weapon Raskolnikov uses is not the one he had planned to use, but a different axe, supplied unexpectedly by the situation.[318] With similar fortuitousness, a hay wagon happens to conceal his entry into the courtyard where the pawnbroker lives.[319] The stairway happens to be empty on his way up, and he finds that the people directly below the pawnbroker happen to have vacated their apartment.[320] Lizaveta happens to arrive just after the murder, prompting Raskolnikov to kill her as well.[321] Two people happen up the stairs as Raskolnikov is readying to escape, and one grows suspicious, leaving the other to guard the apartment while he goes to get the caretaker.[322] But the second man happens to grow impatient and leave his post, giving Raskolnikov just enough time to escape.[323] On his way down the stairs, just before Raskolnikov is about to run into the men who are on their way up to the apartment to catch him, he happens upon an open and empty apartment, which two painters happen to have run out of moments before, "as if by design," allowing him to elude his captors.[324] The situation guides Raskolnikov all the way home, ensuring that, when he arrives, the caretaker of his apartment complex happens to be away, giving Raskolnikov time to return the accidentally found axe to its place.[325]

---

[318] Ibid., 71-72.

[319] Ibid., 73.

[320] Ibid., 73-74.

[321] Ibid., 78-79.

[322] Ibid., 82-84.

[323] Ibid., 84.

[324] Ibid., 84-85.

[325] Ibid., 86.

Perhaps a situationist would want to argue that the situation moved Raskolnikov to commit the crime just as wind bends the daisies. After all, he never would have been willing to go through with the murder, or it would have been impossible, or it would have been only half as bad, or he would not have gotten away with it, were it not for many happenstances that were not in his control and that very well might have been otherwise. Moreover, consistent with the classic situationist interpretation, many of these happenstances are what I described above as "orthogonal" situational features. However, we must bear in mind that no number of happenstances could have driven Raskolnikov to murder if he hadn't formed and reformed the intention to murder in the first place. Raskolnikov had already justified the crime through various acts of existential framing, which we see expressed in his confession to Sonya. Thus, there are really two devils here, one internal and one external, working together to motivate Raskolnikov and help him carry out the crime. In the commission of the crime, Raskolnikov is sometimes bent like a daisy in the wind, but at every stage he is moved by existential framing of the situation, like a daisy following the sun.

It is not only the crime itself that is aided and abetted by situational factors. Raskolnikov's confession and, ultimately, his repentance and redemption might also never have happened were it not for a great many external circumstances that might well have been otherwise. Consider the confession to Sonya. Raskolnikov's powerlessness in confessing repeats the same powerlessness he experienced right before the crime:

> When he reached Kapernaumov's apartment, he felt suddenly powerless and afraid. Thoughtful, he stood outside the door with a strange question: "Need I tell her who killed Lizaveta?" The question was strange because he suddenly felt at the same time that it was impossible not only not to tell her, but even to put the moment off, however briefly. He did not yet know why it was impossible; he only

> felt it, and the tormenting awareness of his powerlessness before necessity almost crushed him.[326]

The moment before Raskolnikov confesses to Sonya, "was terribly like the one when he had stood behind the old woman, having already freed the axe from its loop, and realized that "there was not another moment to lose."[327] Both the form and content of Raskolnikov's confession are dramatically different from what he would have liked them to be. He does not have full agency, and is struck by this from the first moment of confession: "This was not the way, this was not at all the way he had intended to reveal it to her, but thus it came out."[328] He explicitly offers situational features to explain this failure: "I haven't talked with anyone for a long time, Sonya … I have a bad headache now."[329] He vacillates, moved by unexpected turns of mind.[330] We see how physical illness and social isolation affect his thought and speech: "The fever had him wholly in its grip. He was in some sort of gloomy ecstasy. (Indeed, he had not talked with anyone for a very long time!)"[331]

After the confession to Sonya and, later, his confession to the police, Raskolnikov is sent to a prison camp in Siberia. For the first year, he is not repentant but only ashamed that he turned out to be not a Napolean but an ordinary person, unworthy and incapable of carrying through his original plan. However, the angel of the situation is at work on Raskolnikov's behalf. Following

---

[326] Ibid., 406.

[327] Ibid., 409.

[328] Ibid., 411.

[329] Ibid., 416-417.

[330] "'No, Sonya, that's not it!' he began again, suddenly raising his head, as if an unexpected turn of thought had struck him and aroused him anew" (Ibid., 417).

[331] Ibid., 418.

a period of illness and strange dreams in prison, transformed by the new circumstances of his incarceration and by a love for Sonya that emerges altogether unexpectedly, Raskolnikov eventually finds his way to a repentance that is totally organic, totally disconnected from his willful ratiocinations in the city of Petersburg. Just as the situation helped Raskolnikov to commit the crime and get away with it, now the situation also promotes his confession, repentance, and redemption. The situation, Dostoevsky makes clear, is both devil and angel to Raskolnikov.

Conclusion

A sustained engagement with Dostoevsky's work reveals a model of moral psychology that is profoundly different from anything else in the current field. This model emphasizes a moral complexity that arises from the very source of morality, deeper than all values and principles: the intrinsic value we attribute to beings in the world with which we are meaningfully involved. In conversation with many valuable beings, we have many voices in our conscience, a polyphony that changes from moment to moment as our ontologies change—fluctuating with our changing understanding and attention, attachments and moods. Understanding moral psychology from any disciplinary or analytic perspective necessarily requires attention to the ontological roots of morality. These roots are the source of the real polyphonic dynamics of everyday moral cognition.

End Notes

[i] Mediating Factors in Existential Framing

In the above, I have presented existential framing in terms of straightforward subject-object or subject-subject existential relations, for simplicity's sake. The reality is a bit more complex, however, as our normative perception and judgment is also always mediated by systemic relations among objects and concepts. For instance, the best chair of the chairs available to me is probably de facto a "good" chair. But if I had a different array of options, the same chair might be the worst available, and as such it would be a "bad" chair. This means, my normative framing of the chair as good or bad is mediated by the presence or absence of other objects—and the same is true for normative judgments of people.

There are also important systemic dynamics at the level of concepts. Since the posthumously published lectures of Ferdinand de Saussure in the early 20th century, linguists have understood that language is a system, wherein the definition of one word affects the meaning of other words. As an illustration, and sticking with our earlier example, imagine the meaning of the word "chair" in a language where there is no word for "couch." A person who speaks and thinks in such a language, encountering what we call "couch" would probably describe it using a word we might commonly translate as "chair." For them, "chair" is a broader concept than it is for us, a concept that includes couch-like objects. For us, on the other hand, simply having the word "couch" affects the meaning of our word "chair." Specifically, since our language contains a word for couches, our concept of "chair" excludes couch-like objects. And this has consequences for existential framing. Among people with a "couch"-less language, the object we call "couch" might be a perfectly good chair—a surprisingly large and comfy chair, an Uber-chair. Among us, however, a couch—even a good couch—is not a good chair, if what we

really want is a chair. This dynamic is pervasive in language and, therefore, in thought. What something is understood to be—and our normative expectations for how it *ought* to be—is a function not only of straightforward subject-object relations but also of relations among concepts within a language system.

[ii] Politics and Existential Framing

"What white people have to do, is try and find out in their own hearts why it was necessary to have a nigger in the first place, because I'm not a nigger, I'm a man, but if you think I'm a nigger, it means you need it." –James Baldwin, *I Am Not Your Negro*

The most fraught political discussions hinge almost entirely on existential framing. Is the attacker a terrorist or a "radical Islamic" terrorist, or a lone gunman with mental health issues? Are they "illegal aliens" or immigrants? Are they communists or capitalists, socialists or capitalists, liberals or conservatives? At the limit—are they human or *sub*human? Race is an existential frame. Gender is an existential frame. Ideology is an existential frame. Even *species* is an existential frame, which imposes norms upon us. We frame our relations to places. Is earth our mother? Are the rivers and lakes and mountains our relatives? Or is the earth a source of resources? A standing reserve, which we may tap in myriad ways? Something to establish our dominion over? Do we own the land, or are we guests upon the land? Work is an existential frame. Class is an existential frame. Am I upper class or lower class? Wealth can be an existential frame. Am I a one percenter or a ninety-nine percenter? My own aspirations, even utterly secret and unique ones, provide existential framing. Raskolnikov's idea is an existential frame. He *is* the person with his idea. He has a responsibility and a sense of loyalty towards his idea.

All the most fraught debates about facts—ideas about genetic differences between races and sexes, debates about rates of crime, privilege, systemic racism and genderism, access, wage inequality, and the personal, social, and historical causes of these things—are engaged in for the sake of existential framing of ourselves and others. This is equally true for facts that are true and scientifically valid as it is for "alternative facts" and pseudoscience. Racists peddle pseudo-science about deep biological differences between races, in order to buttress their racist existential framing of non-whites as inferior in various ways. Their opponents counter with arguments that buttress their own egalitarian existential framing of the same populations. The arguments of the latter may be perfectly rational, but the reason the facts matter is not simply because they are true but, rather, because they have normative and, ultimately, moral implications.

Existential framing is not an abstract questioning that we participate in as disinterested observers—rather, we research these questions and argue about them precisely to contest answers that already exist, which are already being used to frame people existentially and thereby to support racist or misogynist or egalitarian agendas. Existential framing is the fundamental ground of the psychological reasons why evidence and rational arguments do so little to change people's minds about controversial issues. An existential frame is only loosely connected to facts. Before we know any facts, we have already taken up a normative attitude; we already understand what various groups *are* in this qualitative sense; we are already *in* existential frames. All these facts can be used either to reinforce or destroy the existential frames that already define ourselves and our world. Changing our mind means destroying some important aspect of our identity or our world. Someone else's facts thereby threaten, in the qualified sense of this analysis, our very existence.

[iii] Existential Framing and Empathy

I've already pointed out that empathy has both a rational, "cognitive" component and an extra-rational "affective" component. But empathy is more than this relation between cognitive understanding of the faculties and needs of others and affective compassion for them. My empathy for someone is affected by my esteem for *who they are*, my sense of ownership over them, my intimate connection to them, the goals I share with them, the group-membership I share with them. It has been experimentally shown, for instance, that empathy for one's opponents or enemies is sometimes dampened in the context of inter-group competition (see Cikara et al., "Us and Them: Intergroup Failures of Empathy, 2011). And it is a commonplace to note that we feel greater empathy for friends than for strangers. This means that the affective component of empathy is not only responsive to rational inferences about the mental and dispositional states / features of others, but it can also be turned up or down by existential framing.

[iv] Existential Framing and Heidegger

Key to Heidegger's fundamental ontology is a recognition that all objective scientific knowledge builds upon a prior existential understanding—and that no objective scientific knowledge is meaningful apart from its relation to this existential understanding. Existential understanding is expressed in our ability to navigate our world. We learn how to do things—how to use tools, how to read signs, how to respond to plants, animals, and other people—in view of meaningful activities like building, navigating, communing, and so on. And this practical understanding gives rise to a basic form of interpretation of things, which Heidegger calls the "as" structure of interpretation (39 Being and Time 144, Sein und Zeit 148-149). For instance,

our initial understanding of a hammer is not that it is an object of certain physical proportions, composed of certain materials, or any other objective properties. Indeed, if one were to take the measurements of a hammer without having any idea of what a hammer is for, one would not be interpreting the object as a hammer at all. It might just as well be an ancient weapon, a dildo, a lightning rod, or the eating-utensil of a giant: as a merely objectified object, however, it is none of these things. As we learn how to use a hammer, we gain an implicit understanding of a hammer *as* being for hammering, within a framework where hammering is for building and building is for any number of things. The most basic form of interpretation of things in the world is simply an explicit working-out of this "as" structure, which is already implicit in understanding (an understanding that we express simply in skillfully doing things—Heidegger calls this implicit knowing-how relation to things the readiness-to-hand of things). The activity of hammering is similarly interpreted *as* something for building or tearing down. And building and tearing down are projects that have meaning in relation to the needs of the moment, our larger goals, the people we are helping or hindering, and so on. Thus, we automatically interpret everything in terms of a meaningful context that is always in view—most of the time, only implicitly.

Heidegger suggests that it is a mistake to assume that the essence of things is in their objective properties. In fact, the essence of any thing is extended in a kind of *subjective* time and space. The essence of a hammer—its fundamental meaning—is determined both by the general "what" of what hammers are used for and by more specific considerations of the use a particular hammer is fitted or destined for. Its essence is what it means to us—and this meaning is extended across the "space" of our attachments, involvements, and complex projects and the "time" of our

past-, present-, and future-oriented activities. It is a complete mistake to imagine that any objective measurement of any hammer has meaning apart from this interpretive structure.

Moral psychology has a similar problem to that of objectivistic ontology. We are continually trying to understand what moral cognition consists of—What is the deeper grammatical structure of moral reasoning? Do moral principles or situational factors matter more? Does domain-general rational processing or domain-specific processing play a bigger role?—but we have failed to take account of the interpretive structure that organizes moral judgment in every situation. Moral judgment of someone is always a judgment of someone *as* a particular someone. This as-structure of a person's identity precedes and incorporates any objective measure of, say, their biological properties. I am my mother's son, my sister's brother, my country's citizen, a member of my church or university, a person with certain ambitions, fears, and ideals—and I am also someone with the cognitive capacities to understand moral rules, to empathize with other people, to appreciate the difference between a good person and a bad person, and so on. All these features feed into others' understanding of who and what I am. All these features help others predict how I am likely to behave. All these features contribute to assumptions about how I ought to be. However, the relevance of this existential specificity of our understanding of existence for moral cognition has so far eluded psychological theories that focus on objective features of individuals and groups.

[v] Equivocality of the "Existential Framing" Concept

There are a couple potential equivocations in the language I have been using to talk about existential framing. On one hand, I have talked about how existential framing *determines* or reciprocally co-defines ourselves and the things with which we are meaningfully involved. On

the other hand, this process has been described as an *interpretive* framing process. The obvious question this raises is whether I am talking about a second-order process of interpreting something that already exists independently or a first-order process of constituting something out of whole cloth. However, I don't actually want to indulge in any philosophical argument over whether things *really* exist independent of our interpretation of them, and I don't have to. I will simply acknowledge that the objective existence of things is a useful and ubiquitous assumption that people make most of the time, until perhaps we take the right drugs—and this common-sense idea that things have objective existence is an assumption I make in this work. At the same time, I also acknowledge what we have known since at least Kant: that every designation of what something is is necessarily an interpretation constructed by our minds, so that we can never actually encounter the objective "thing in itself" outside of our interpretive perspective. Thus, it is impossible to disentangle interpretation from determination in the dynamic of existential framing. Existential framing is both an interpretive framing of a thing and a determinative constitution of a thing. This does *not* mean that existential framing determines the objective properties of a thing—throughout this work, I *assume* that things have objective properties that are completely independent of interpretation or perception. It simply means that, if I speak of the meaning of a hammer as being-for-hammering, this understanding is both interpretive and constitutive of its existential essence. Or if I identify myself as my mother's son, this is both an interpretation of who I am and a constitution of myself as this thing—and to say that it is existentially *constitutive* is in no way to imply that, for instance, biological relatedness isn't real or doesn't matter.

There is another way in which existential framing is an equivocal concept, which has implications for attempts to experimentally test it. Let's say I want to distinguish people's

normativity of perception, judgment, and action towards a plant that is understood as a weed vs. a plant that is understood as a beautiful flower. People might relate to a "weed" differently than a "flower" simply because the name is itself value-laden, with flowers being more highly valued than weeds. And this difference would be attributable to existential framing. However, each of these words might also conjure a factually different state of affairs. Calling it a "weed," I might imply that this plant is an invasive species that farmers and gardeners have to actively resist, for the sake of other plants; calling it a "flower," I might imply that this plant is a species that farmers and gardeners actively cultivate for its beauty and scent. This expression of facts is *also* existential framing.

What this shows is that existential framing might be either empirical or extra-empirical, rational or extra-rational, though it is most often a bit of both. In the examples above, the former case is a kind of extra-empirical, extra-rational aspect of existential framing, whereas the latter case shows how a simple understanding of empirical facts constitutes an existential frame. While it is interesting that existential framing is sometimes unmoored from facts, we should recognize that facts are powerful existential frames too. If we fail to appreciate this, we will underestimate existential framing. We will see only the ideologically determined part. In experiments, we will tend to match stimuli on factual information and measure only this extra-rational surplus. Doing so, we will find effects, but they will always be an underestimation of the role of existential framing. Experimentally demonstrating the true importance of existential framing, on the other hand, might be very difficult.

Bibliography

The Perfect Generosity of Prince Vessantara. Translated by Margaret Cone and Richard Gombrich. Oxford: Clarendon Press, 1977.

Alberts, Susan, Jeanne Altmann, Diane Brockman, Marina Cords, Linda Fedigan, Anne Pusey, Tara Stoinski, Karen Strier, William Morris, and Anne Bronikowski. "Reproductive Aging Patterns in Primates Reveal that Humans are Distinct." *Proceedings of the National Academy of Sciences of the United States of America* 110.33 (2013): 13440–13445.

Appiah, Kwame. *Experiments in Ethics*. Cambridge: Harvard University Press, 2008.

Aquinas, Thomas, *Summa Theologica*. Translated by Fathers of the English Dominican Province. Ohio: Benziger Bros. Edition, 1947. Digital File.

Ambady, Nalini, Debi Laplante, Thai Nguyen, Robert Rosenthal, Nigel Chaumeton, and Wendy Levinson. "Surgeons' Tone of Voice: A Clue to Malpractice History." *Surgery* 132 (2002): 5-9.

Aristotle. *The Nicomachean Ethics.* Edited by E. Capps, T.E. Page, and W.H.D. Rouse and Translated by H. Rackham. London: William Heineman, 1934.

Bakhtin, Mikhail. *Problems of Dostoevsky's Poetics.* Edited and Translated by Caryl Emerson. Minneapolis: University of Minnesota Press, 1999. Kindle file.

Bar, Moshe, Maital Neta, and Heather Linz. "Very First Impressions." *Emotion* 6.2 (2006): 269-278.

Batson, Daniel. *The Altruism Question: Toward a Social-Psychological Answer*. NJ: Lawrence Erlbaum, 1991.

Bellah, Robert, Richard Madsen, William Sullivan, Ann Swindler, and Steven Tipton. *Habits Of the Heart: Individualism and Commitment in American Life.* Berkeley: University of California Press, 1985.

Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation*. White Dog Publishing, 2010 (originally published in 1781), Kindle Book.

Bigelow, Ann and Philippe Rochat. "Two-Month-Old Infants' Sensitivity to Social Contingency in Mother–Infant and Stranger–Infant Interaction." *Infancy* 9.3 (2006): 313-325.

Carman, Taylor. *Heidegger's Analytic: Interpretation, Discourse, and Authenticity in* Being and Time. Cambridge: Cambridge University Press, 2003.

Cikara, Mina, Emile Bruneau, and Rebecca Saxe. "Us and Them: Intergroup Failures of Empathy." *Current Directions in Psychological Science* 20.3 (2011): 149-153.

Clement, Fabrice, Stephane Bernard, and Laurence Kaufmann. "Social Cognition is not Reducible to Theory of Mind: When Children Use Deontic Rules to Predict the Behaviour of Others." *British Journal of Developmental Psychology* 29 (2011): 910-928.

Croft, Darren, Lauren Brent, Daniel Franks, and Michael Cant. "The Evolution of Prolonged Life After Reproduction, *Trends in Ecology & Evolution* 30.7 (2015): 407 – 416.

Damasio, Antonio. *Descartes' Error: Emotions, Reason, and the Human Brain*. New York: Avon Books, 1994.

Darwin, Charles. *The Descent of Man and Selection in Relation to Sex.* London: John Murray, 1871. Volume II

Darwin, Charles. *On the Origin of Species by Means of Natural Selection*: *Or the Preservation of Favored Races in the Struggle for Life*. American Edition. New York: D. Appleton & Company, 1861.

Darwin, Francis ed. *The Life and Letters of Charles Darwin, Including an Autobiographical Chapter*. London: John Murray, 1887. Volume II.

Dawkins, Richard. "Twelve Misunderstandings of Kin Selection." *Zeitschrift für Tierpsychologie* 51 (1979): 184-200.

DeSteno, David, Cynthia Breazeal, Robert Frank, David Pizarro, Jolie Baumann, Leah Dickens, and Jin Joo Lee. "Detecting the Trustworthiness of Novel Partners in Economic Exchange." *Psychological Science* 23 (2012): 1549 –1556.

de Waal, Frans. *The Age of Empathy: Nature's Lessons for a Kinder Society*. New York: Harmony Books, 2009.

de Waal, Frans. "Putting the Altruism Back in Altruism: The Evolution of Empathy." *Annual Review of Psychology* 59 (2008): 279-300.

de Waal, Frans, Robert Wright, Christine Korsgaard, Philip Kitcher, and Peter Singer. *Primates and Philosophers.* New Jersey: Princeton University Press, 2006.

Ding, Xiao Pan, Henry Wellman, Yu Wang, Genyue Fu, and Kang Lee. "Theory of Mind Training Causes Honest Young Children to Lie." *Psychological Science* 26.11 (2015): 1812–1821.

Doris, John. *Lack of Character: Personality and Moral Behavior.* New York: Cambridge University Press, 2002.

Dostoevsky, Fyodor. *The Brothers Karamazov: A Novel in Four Parts With Epilogue*. Translated by Richard Pevear and Larissa Volokhonsky. New York: Farrar, Straus and Giroux, 2002. Kindle Edition.

———. *Crime and Punishment.* Translated by Richard Pevear and Larissa Volokhonsky. New York: Vintage Classics, 2012. Kindle file.

———. *The Idiot*. Translated by Richard Pevear and Larissa Volokhonsky. New York: Vintage Books, 2003.

———. *Notes from Underground*. Translated by Jane Kentish. New York: Oxford University Press, 1991.

———. *Demons*. Translated by Richard Pevear and Larissa Volokhonsky. New York: Alfred A. Knopf, Inc., 2005.

———. *A Writer's Diary: Volume One.* Translated by Kenneth Lantz. Evanston: Northwestern University Press, 1993.

———. *A Writer's Diary: Volume Two.* Translated by Kenneth Lantz. Evanston: Northwestern University Press, 1994.

Dunbar, Robin. "Gossip in Evolutionary Perspective." *Review of General Psychology* 8.2 (2004): 100-110.

Durkheim, Emile. *Emile Durkheim on Morality and Society: Selected Writings*. Edited by Robert Bellah. Chicago: The University of Chicago Press, 1973.

Dziobek, Isabel, Kimberly Rogers, Stefan Fleck , Markus Bahnemann, Hauke Heekeren, Oliver Wolf, and Antonio Convit. "Dissociation of Cognitive and Emotional Empathy in Adults with Asperger Syndrome Using the Multifaceted Empathy Test (MET)." *Journal of Autism and Developmental Disorders* 38 (2008): 464–73.

Einstein, Albert. *Relativity: The Special and the General Theory*. Translated by Robert W. Lawson. New York: Barnes & Noble, 2008 (original 1920).

Fisher, Ronald. *The Genetical Theory of Natural Selection.* London: Oxford University Press, 1930.

Gilligan, Carol. *In a Different Voice: Psychological Theory and Women's Development*. Cambridge: Harvard University Press, 1982.

Gintis, Herbert, Joseph Henrich, Samuel Bowles, Robert Boyd, and Ernts Fehr. "Strong Reciprocity and the Roots of Human Morality." *Social Justice Research* 21 (2008): 241-253.

Gintis, Herbert. "Strong Reciprocity and Human Sociality." *Journal of Theoretical Biology* 206 (2000): 169-179.

Goodwin, Geoffrey, Jared Piazza, and Paul Rozin. "Moral Character Predominates in Person Perception and Evaluation." *Journal of Personality and Social Psychology*, 106.1 (2014): 148-168.

Gould, Stephen Jay, and Elisabeth S. Vrba. "Exaptation—A Missing Term in the Science of Form." *Paleobiology*. 8.1 (1982): 4-15.

Gould, Stephen Jay, and Richard C. Lewontin. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme." *Proceedings of the Royal Society of London*. 205.1161 (1979): 581-598.

Graham, Jesse, Jonathan Haidt, Matt Motyl, Peter Meindl, Carol Iskiwitch, and Marlon Mooijman. "Moral Foundations Theory: On the Advantages of Moral Pluralism Over Moral Monism." In *The Atlas of Moral Psychology: Mapping Good and Evil in the Mind.* Edited by Kurt Gray and Jesse Graham. New York: Guilford, in press.

Graham, Jesse, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean Wojcik, and Peter Ditto. "Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism." *Advances in Experimental Social Psychology* 47 (2013): 55-130.

Gray, Kurt, Chelsea Schein, and Adrian Ward. "The Myth of Harmless Wrongs in Moral Cognition: Automatic Dyadic Completion From Sin to Suffering." *Journal of Experimental Psychology: General* Advance Online Publication (2014).

Greene, Joshua, Brian Sommerville, Leigh Nystrom, John Darley, and Jonathan Cohen. "An fMRI Study of Emotional Engagement in Moral Judgment." *Science* 293 (2001): 2105-2108.

Greene, Joshua, Leigh Nystrom, Andrew Engell, John Darley, and Jonathan Cohen. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron* 44 (2004): 389–400.

Haidt, Jonathan. "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108 (2001): 814-834.

———. *The Righteous Mind: Why Good People are Divided by Politics and Religion*. New York: Pantheon. Knopf Doubleday Publishing Group, 2012. Kindle file.

Haidt, Jonathan and Fredrik Bjorklund. "Social Intuitionists Answer Six Questions About Morality." In *Moral Psychology*, *Vol. 2: The Cognitive Science of Morality*, 181-217. Edited by Walter Sinnott-Armstrong. Cambridge: MIT Press, 2008.

Haidt, Jonathan and Joseph, Craig. "The Moral Mind: How 5 Sets of Innate Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules". In *The Innate Mind, Vol. 3*. Edited by P. Carruthers, S. Laurence and S. Stich. New York: Oxford, 2007, 367-391.

Haidt, Jonathan, Silvia Koller, and Maria Dias. "Affect, Culture and Morality, or is it Wrong to Eat Your Dog?" *Journal of Personality and Social Psychology* 65 (1993): 613-628.

Hamlin, Kiley. "The Case for Social Evaluation in Preverbal Infants: Gazing Toward One's Goal Drives Infants' Preferences for Helpers over Hinderers in the Hill Paradigm." *Frontiers in Psychology* 5 (2014a): 1563. DOI:10.3389/fpsyg.2014.01563

Hamlin, Kiley. "Context-Dependent Social Evaluation in 4.5-Month-Old Human Infants: The Role of Domain-General Versus Domain-Specific Processes in the Development of Social Evaluation." *Frontiers in Psychology* 5 (2014b): 614. DOI: 10.3389/fpsyg.2014.00614

Hamlin, Kiley, Karen Wynn, Paul Bloom, and Neha Mahajan. "How Infants and Toddlers React to Antisocial Others." *PNAS* 108.50 (2011): 19931-19936. DOI: 10.1073/pnas.1110306108

Hamlin, Kiley, Karen Wynn, and Paul Bloom. "Social Evaluation by Preverbal Infants." *Nature* 450 (2007): 557–559.

Hamilton, William. "The Genetical Theory of Social Behaviour." *Journal of Theoretical Biology* 7 (1964): 1-52.

Haney, Craig, Curtis Banks, and Philip Zimbardo. "Interpersonal Dynamics in a Simulated Prison." *International Journal of Criminology and Penology* 1 (1973): 69-97.

Hawkes, Kristin, James O'Connell, Nicholas Blurton Jones, Helen Alvarez, and Eric Charnov. "Grandmothering, Menopause, and the Evolution of Human Life-Histories. Proceedings of the National Academy of Sciences U.S.A. 95 (1998): 1336–1339.

Hawkes, Kristen and Rebecca Bliege Bird. "Showing off, Handicap Signaling, and the Evolution of Men's Work." *Evolutionary Anthropology* 11 (2002): 58-67.

Heidegger, Martin. *Being and Time*. Translated by Joan Stambaugh and Dennis Schmidt. Albany: State University of New York Press, 2010 [original 1927].

Heidegger, Martin. *Zein und Zeit*. Tubingen: Max Niemeyer Verlag, 2006 [original 1927].

Heiphetz, Larissa, Nina Strohminger, and Liane Young. "The Role of Moral Beliefs, Memories, and Preferences in Representations of Identity." *Cognitive Science* 41 (2016): 744-767.

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, Frank Marlowe, David Tracer, and John Ziker. "Costly Punishment Across Human Societies." *Science* 312.5781 (2006): 1767-1770.

Henrich, Joseph and Francisco Gil-White. "The Evolution of Prestige: Freely Conferred Deference as a Mechanism for Enhancing the Benefits of Cultural Transmission." *Evolution and Human Behavior* 22.3 (2001): 165–96.

Henrich, Natalie and Joseph Henrich. *Why Humans Cooperate: A Cultural and Evolutionary Explanation*. Oxford: Oxford University Press, 2007.

Hoffman, Martin. *Empathy and Moral Development*. Cambridge: Cambridge University Press, 2000.

Hoffman, Martin. "Developmental Synthesis of Affect and Cognition and its Implications for Altruistic Motivation." *Developmental Psychology* 11 (1975): 607–622.

Huebner, Bryce, James Lee, and Marc Hauser. "The Moral-conventional Distinction in Mature Moral Competence." *Journal of Cognition and Culture* 10.1-2 (2010): 1-26.

Kant, Immanuel. *Critique of the Power of Judgment.* Edited by Paul Guyer. Translated by Paul Guyer and Eric Matthews. New York: Cambridge University Press, 2000.

Kant, Immanuel. *Groundwork for the Metaphysics of Morals*. Edited and Translated by Allen Wood. New Haven: Yale University Press, 2002 (originally published in 1785).

Kaplan, Hillard, Paul Hooper, and Michael Gurven. "The Evolutionary and Ecological Roots of Human Social Organization." *Philosophical Transactions of the Royal Society B* 364 (2009): 3289-3299.

Kiehl, Kent and Morris Hoffman. "The Criminal Psychopath: History, Neuroscience, Treatment, and Economics." *Jurimetrics* 51 (2011): 355–397.

Kohlberg, Lawrence. *The Philosophy of Moral Development: Moral Stages and the Idea of Justice*. Vol. 1. New York: Harper & Row, 1981.

Kohlberg, Lawrence. *The Psychology of Moral Development: The Nature and Validity of Moral Stages*. Vol. 2. New York: Harper & Row, 1984.

Mayr, Enrst. "Cause and Effect in Biology." *Science* 134.3489 (1961): 1501-1506.

Meffert, Harma, Valeria Gazzola, Johan den Boer, Arnold Bartels, and Christian Keysers. "Reduced Spontaneous but Relatively Normal Deliberate Vicarious Representations in Psychopathy." *Brain* 136.8 (2013): 2550-2562.

Mikhail, John. "Universal Moral Grammar: Theory, Evidence and the Future." *TRENDS in Cognitive Sciences* 11.4 (2007): 143-152.

Mill, John Stuart. *Utilitarianism*. Kitchener: Batoche Books, 2014 [originally published in 1863].

Miller, Geoffrey. "Sexual Selection for Moral Virtues." *The Quarterly Review of Biology* 82.2 (2007): 97-123.

Montgomery, Charlotte, Carrie Allison, Meng-Chuan Lai, Sarah Cassidy, Peter Langdon, and Simon Baron-Cohen "Do Adults with High Functioning Autism or Asperger Syndrome Differ in Empathy and Emotion Recognition*?" Journal of Autism and Developmental Disorders* 46 (2016):1931-1940.

Moran, Joseph, Liane Young, Rebecca Saxe, Su Mei Lee, Daniel O'Young, Penelope Mavros, and John Gabrieli. "Impaired Theory of Mind for Moral Judgment in High-Functioning Autism." *Proceedings of the National Academy of Sciences* 108.7 (2011): 2688–2692.

Nietzsche, Friedrich. *Beyond Good and Evil.* Translated by Helen Zimmern. New York: Barnes and Noble, 2007.

Nowak, Martin and Karl Sigmund. "Evolution of Indirect Reciprocity." *Nature* 437.27 (2005): 1291-1298.

Patil, Indrajeet and Giorgia Silani. "Reduced Empathic Concern Leads to Utilitarian Moral Judgments in Trait Alexithymia." *Frontiers in Psychology* 5.501 (2014): 1-12.

Phillips, Jonathan and Fiery Cushman. "Morality Constrains the Default Representation of What is Possible." *Proceedings of the National Academy of Sciences* 114.18 (2017): 4649-4654.

Piaget, Jean. *The Moral Judgment of the Child*. Translated by Marjory Gabain. New York: The Free Press, 1965 [originally published in 1948].

Pizarro, David and David Tannenbaum. "Bringing Character Back: How the Motivation to Evaluate Character Influences Judgments of Moral Blame." In *The Social Psychology of Morality: Exploring the Causes of Good and Evil*, 91-108. Edited by Mario Mikulincer and Phillip Shaver. Washington, DC: APA Press, 2011.

Pizarro, David, Erich Uhlmann, and Peter Salovey. "Asymmetry in Judgments of Moral Blame and Praise: The Role of Perceived Metadesires." Psychological Science 14 (2003): 267-272.

Plato. *'Protagoras' and 'Meno'*. Translated by Robert Bartlett. Ithaca: Cornell University Press, 2004.

Premack, David, and Guy Woodruff. "Does the Chimpanzee Have a 'Theory of Mind'?" *Behavioral and Brain Sciences* 4 (1978): 515-526.

Reicher, Stephen and Alexander Haslam. "After Shock? Towards a Social Identity Explanation of the Milgram 'Obedience' Studies." *British Journal of Social Psychology* 50 (2011): 163-169.

Richerson, Peter and Robert Boyd. *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press, 2005. Kindle file.

Robbins, Erin and Philippe Rochat. "Emerging Signs of Strong Reciprocity in Human Ontogeny." *Frontier in Psychology* 2.353 (2011).

Rochat, Philippe, Maria Dias, Guo Liping, Tanya Broesch, Claudia Passos-Ferreira, and Ashley Winning. "Fairness in Distributive Justice by 3- and 5-year-olds Across Seven Cultures." *Journal of Cross Cultural Psychology* 40.3 (2009): 416–442.

Rochat, Philippe. "Layers of Awareness in Development." *Developmental Review* 38 (2015): 122-145.

Rosenstein, Diana and Harriet Oster. "Differential Facial Responses to Four Basic Tastes in Newborns." *Child development* 59.6 (1988): 1555-1568.

Rottman, Joshua, Deborah Keleman, and Liane Young. "Purity Matters More than Harm in Moral Judgments of Suicide." *Cognition* 133.1 (2014): 332-334.

Royzman, Edward, Kwanwoo Kim, and Robert Leeman. "The Curious Tale of Julie and Mark: Unraveling the Moral Dumbfounding Effect." *Judgment and Decision Making* 10.4 (2015): 296-313.

Ryan, Michael. "Sexual Selection, Sensory Systems, and Sensory Exploitation." *Oxford Surveys in Evolutionary Biology*. Edited by Douglas Futuyama & Janis Antonovics. London: Oxford University Press, 1990. Volume 7: 157–195.

Ryan, Michael, James Fox, Walter Wilczynski, and A. Stanley Rand. "Sexual Selection for Sensory Exploitation in the Frog Physalaemus Pustulosus." *Nature* 343 (1990): 66-67.

Sanfey, Alan, James Rilling, Jessica Aronson, Leigh Nystrom, and Jonathan Cohen. "The Neural Basis of Economic Decision-making in the Ultimatum Game." *Science* 300 (2003): 1755-1758.

Scarf, Damien, Kana Imuta, Michael Colombo, and Harlene Hayne. "Social Evaluation or Simple Association? Simple Associations May Explain Moral Reasoning in Infants." *PLoS ONE* 7.8 (2012).

Schein, Chelsea and Kurt Gray. "The Unifying Moral Dyad: Liberals and Conservatives Share the Same Harm-Based Moral Template." *Personality and Social Psychology Bulletin* 41 (2015): 1147–1163.

Simner, Marvin. "Newborn's Response to the Cry of Another Infant." *Developmental Psychology 5* (1971): 136-150.

Smith, Adam. *The Theory of Moral Sentiments*. Oxford: Oxford University Press, 1976 (Originally published in 1759). Kindle File.

Smith, Christian. *Moral, Believing Animals: Human Personhood and Culture*. Oxford: Oxford University Press, 2003.

Sober, Elliot and David Sloan Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge: Harvard University Press, 1998.

Strohminger, Nina and Shaun Nichols. "Neurodegeneration and Identity." *Psychological Science*, 26.9 (2015): 1469-1479.

Strohminger, Nina and Shaun Nichols. "The Essential Moral Self." *Cognition*, *131* (2014): 159–171.

Talwar, Victoria and Kang Lee. "Social and Cognitive Correlates of Children's Lying Behavior." *Child Development* 79.4 (2008): 866-881.

Taylor, Charles. *Modern Social Imaginaries*. Durham: Duke University Press, 2004.

Tinbergen, Nikolaas. "On the Aims and Methods of Ethology." *Zeitschrift für Tierpsychologie* 20 (1963): 410–433.

Trivers, Robert. "The Evolution of Reciprocal Altruism." *The Quarterly Review of Biology* 46 (1971): 35–57.

Turiel, Elliot. *The Development of Social Knowledge: Morality and Convention*. Cambridge, U.K.: Cambridge University Press, 1983.

Uhlmann, Erich, David Pizarro, and Daniel Deirmeier. "A Person-centered Approach to Moral Judgment." *Perspectives on Psychological Science* 10 (2015): 72-81.

Uhlmann, Erich and Luke Lei Zhu. "Acts, Persons, and Intuitions: Person-centered Cues and Gut Reactions to Harmless Transgressions." *Social Psychological and Personality Science* 5 (2014): 279–285.

Uhlmann, Erich, Luke Lei Zhu, and Daniel Diermeier. "When Actions Speak Volumes: The Role of Inferences About Moral Character in Outrage Over Racial Bigotry." *European Journal of Social Psychology* 44 (2014): 23–29.

Westra, Evan. "Character and Theory of Mind: An Integrative Approach." *Philosophical Studies* (Advance Publication: 2017).

Williams, George C. *Adaptation and Natural Selection*. Princeton: Princeton University Press, 1966 (8th Edition, 1996).

Wimmer, Heinz, and Josef Perner. "Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception." *Cognition* 13 (1983): 103-128.

Wynne-Edwards, Vero Copner. *Animal Dispersion in Relation to Social Behavior.* Edinburgh: Oliver and Boyd, 1962.

Zahavi, Amotz. "Altruism as a Handicap: the Limitations of Kin Selection and Reciprocity." *Journal of Avian Biology* 26 (1995): 1–3.

Zahavi, Amotz. "Mate Selection – A Selection for a Handicap." *Journal of Theoretical Biology* 53 (1975): 205–214.

Zahn-Waxler, Carolyn, Marian Radke-Yarrow, and Elizabeth Wagner. "Development of Concern for Others." *Developmental Psychology* 28.1 (1992): 126-136.