INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

IMMORAL PSYCHOLOGY

THE COGNITIVIST'S CONUNDRUM

by

JOSEPH STEPHEN BIEHL

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2003

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

1/29/03
Date

Chair of Examining Committee

1/29/03
Date

Executive Officer

Stefan Baumrin

Simon Blackburn

Steven Cahn
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

IMMORAL PSYCHOLOGY

The Cognitivist's Conundrum

By

Joseph Stephen Biehl

Adviser: Professor Stefan Baumrin

That people do wrong would appear to be a *moral* datum: a moral realm without wrongdoing may not be coherent. Thus, an adequate philosophic theory of morality ought to allow for it. But such a theory ought also to explain wrongdoing, both axiologically and causally. This is so if we take such a theory to have practical significance. Indeed, insofar as moral philosophy and its cognate areas have practical significance, explaining wrongdoing is arguably the most pressing practical issue for theory construction in this domain. For a philosophical theory of morality to have practical significance, and, further, to have influence on behavior, it must be psychologically relevant to the organisms that it ranges over. That is, the concepts that the theory purports to explain must be plausibly realized within the psychology of the creatures for which the theory is intended. Any theory satisfying this constraint would then be in a position to illuminate those psychological features that are explanatory of blameworthy behavior. The argument presented here is that moral realism and its underlying moral psychology, cognitivism, face serious difficulties allowing for and explaining, both metaphysically and psychologically, such behavior. Moral cognitivism and realism fail to adequately account for this most fundamental of moral phenomena. Hence, a comprehensive understanding of moral experience is impossible within the cognitivist/realist perspective.

Acknowledgments

What a long, strange trip it's been. Much has happened and much has changed since this journey began back in the autumn of '91. And if I have only myself to blame for how long it took to get here I nevertheless have many to thank for making getting here possible. First and foremost I must thank my mother and father, brothers and sister. Without their unending and unwavering support (material, emotional, and spiritual) my pursuit of a doctorate in philosophy may well have been abandoned during any one of the long, dark nights my soul has seen these past eleven (*eleven!*) years. My debt to you can never be repaid. I only wish Grandma and Grandpa could be reading this with you.

Clearly this dissertation could not have been written (and finished!) without the support, guidance, generosity, and wisdom of the many professors who have helped me along the way. Charles Landesman's welcoming and kind nature served me well from the beginning to the end of my graduate career, and made one of the more difficult decisions for me considerably easier. My gratitude for the way Jerry Fodor would invite a raw, embarrassingly ignorant young graduate student into his office for lengthy philosophical discussions, and for how he endured at times equally extensive phone conversations, is endless. He was the greatest private (and unpaid!) tutor a philosophy student could possibly have. And to Stephen Schiffer, who could not imagine how encouraging he was to me when he told me that philosophers are like tennis players: they're not born, they're made. David Rosenthal, to me, has been a model for how a graduate faculty member can challenge, encourage, enrich, and befriend those who's dream is to become what he is: a philosopher.

Table of Contents

And then there is my dissertation committee, a group that steered, cajoled, suffered, and ultimately, welcomed me to my final destination. I owe them so much. Patricia Smith's support and clarity of vision has been invaluable. She is also the best Chairperson an aspiring teacher could possibly have. John Greenwood has treated me more like a colleague and friend than a student and his wisdom, humor, gentleness has always left me feeling better leaving his office than when I entered it. Simon Blackburn's generosity and kindness knows no bounds and he has provided me, through his teaching and his work, the greatest inspiration. Without Steven Cahn's wisdom I would still be writing my dissertation. He never lost sight of the end result even though I did several times. He, too, is an exemplar for how a faculty member should handle the apprentice. And lastly, and mostly, I must thank my advisor, Stefan Baumrin. He deserves a medal for his patience with my methods, petulance, thin skin, and arrogance. He oversaw not only the successful completion of my dissertation but my reluctant (and, sadly, only partial) maturation. He made me a much better philosopher and I owe him a great deal. I only hope that I have not hastened his retirement. Thank you, Stefan.

Graduate school would not have been worthwhile without the friends I made there. They were my classmates, cohorts, collaborators, and chums. Four that deserve special mention are Daniel Kaufman, Martin Harvey, David Shein and Gerrit Jan Kamperdyk. They provided me not only with their friendship but with encouragement, advice, enlightenment, support, criticism, and, best of all, memories. Our discussions and debates, outside of school by day and in the pubs at night, have been the best part my experience. You will be remembered fondly forever.

Finally there is Marisa. Your love, faith, patience, and charity enabled me to reach my goal, for you are worth finishing for. You are the love of my life and I look forward to spending the rest of it standing by your side the way you have stood by me. Thank you for believing in me. I love you.

P.S. *Now* I can quit smoking!

J.S.B. New York, 2003

1

Explaining Wrongdoing

> ...It is well said, then, that it is by doing just acts that the just man is produced, and by doing temperate acts the temperate man; without doing these no one would have even a prospect of becoming good.
>
> But most people do not do these, but take refuge in theory and think they are being philosophers and will become good in this way, behaving somewhat like patients who listen attentively to their doctors, but do none of the things they are ordered to do. As the latter will not be made well in body by such a course of treatment, the former will not be made well in spirit by such a course of philosophy.
>
> —Aristotle, *Nicomachean Ethics*[1]

Arguably, as Professor Moore claimed, "What is good?" and "What is bad?" are the fundamental questions that an ethical theory should aim to address.[2] But there are, as Moore himself acknowledged, many other questions of interest for any ethicist. One of those questions, undoubtedly, is "Why do we do wrong?" This question is of considerable theoretical interest for if we accept as a datum, as surely we must, that moral agents *can* and *do* wrong, the construction of our ethical theories is then constrained both to allow and to explain this fact.[3] But more than this, if as *ethicists* we think that it is part of our task to provide some form of guidance to moral agents the question of why we do wrong becomes particularly pressing. This is because the question "Why do we do wrong?" has even greater practical significance than theoretical. For, as *moral practitioners*, the question may matter more to us than "What is good?" Many, perhaps most, moral agents are (some would say

---

[1] II 4 1105b9-19. All references to the Nicomachean Ethics (hereafter *NE*) are from the W. D. Ross translation in *The Basic Works of Aristotle*, ed. Richard McKeon (New York: Random House, 1941).

[2] G. E. Moore, *Principia Ethica* (1903), Revised Edition, edited by Thomas Baldwin (Cambridge: Cambridge University Press, 1993): 55.

[3] Ethical theories are constrained to allow and explain many other things as well.

falsely) confident of their understanding of what is good and bad and their ability to determine those things that are so identified.[4] Yet in spite of such claims of moral wisdom, wrongdoing abounds. No one has doubts that other people do wrong, yet the reproachful eye of conscience sees little rest itself: self-condemnation is the career of many. Nor can we wonder at our concern to avoid, prevent, or at any rate minimize the wrongdoing that does occur, whomever the source. But this calls for understanding, both of the nature of such behavior and of those capable of engaging it (i.e. all of us). Without such an understanding the measures we would take will most likely be in vain.

It is perhaps surprising that when we consider the practical issue of how unavoidable—and important—the task of providing explanations of wrongdoing is we may begin to have a greater appreciation of the theoretical significance of the issue. It would not be absurd to think that the problem of wrongdoing is really what gives impetus to ethics as a theoretical enterprise. If no one did wrong, if no one ever had a problem with their own or another's behavior, what need would we have for theories of what is right or good?[5] As Kant said, a moral law that takes the form of a *constraint* on the will (as categorical imperatives do) is necessary only for a will that is not good, that is for a will that does not always do what it ought to do, let alone because it ought to do it.[6] Mill understood the purpose of a moral

---

[4] A point that was perhaps most eloquently put by Thomas Hobbes:

> And as for the faculties of mind...I find yet a greater equality amongst men, than that of strength. For Prudence, is but Experience; which equall time, equally bestowes on all men, in those things they equally apply themselves unto. That which may perhaps make such equality incredible, is but a vain conceipt of ones owne wisdome, which almost all men think they have in greater degree, than the Vulgar; that is all men but themselves, and a few others, whom by Fame, or of concurring with themselves, they approve...For there is not ordinarily a greater signe of the equall distribution of any thing, than that every man is contented with his share.

Leviathan (1651), (London: Penguin Books, 1985): Ch. XIII, 183-4.

[5] Surely there would be no practical need.

[6] See his *Foundations of the Metaphysics of Morals* (1785), translated by Lewis White Beck, Indianapolis: Bobbs-Merrill, 1959: 29/413.

theory to tell us unequivocally what our duties are and by what test we may know them.[7] There would be no call for such a service if we were invariably doing our duty to begin with. Without wrongdoing, moral theory would hold only academic interest. It would exist only as a theoretical luxury as opposed to a practical necessity. Explaining wrongdoing, then, provides moral theory not only with practical significance but also the opportunity to comprehend the conditions of its own utility. In accounting for why agents do wrong, ethicists reveal the very presuppositions of their vocation.

Morality is often said to be essentially practical. Taking the utility of moral theory to be dependent upon the actual conditions of our moral practice is one way of making sense of that idea. But if this is so, then in an important sense moral theories are significantly beholden to that practice. We should then expect to find that moral theories aim to augment, explicate, and facilitate the pre-philosophic understanding of morality that informs our moral practice. And, indeed, this is what we do find. Kant, for instance, aimed to reveal the principles that would make our commonsense intuitions concerning morality *possible.* The moral theorist's *modus operandi* is to make sense of morality as she finds it. This is most plainly in evidence when we consider the conceptual repertoire that the moral theorist employs. The vocabulary of moral theory is that of the everyday, whether the theory aims to tell us what we ought to do and who we ought to be, or what such a notion as 'ought' is supposed to mean. The elements of moral vocabulary are not terms of art, but are the common property of the vulgar and the wise. This suggests that the proper attitude of the theorist towards commonsense is, essentially, respectfully conservative. Aristotle put it thus:

> We must, as in all other cases, set the observed facts before us and, after first discussing the difficulties, go on to prove, if possible, the truth of all the common opinions about these affections of the mind, or, failing this, of the greater number

---

[7] J.S. Mill, *Utilitarianism* (1861/63), edited by Roger Crisp (Oxford: Oxford University Press, 1998): 65.

> and the most authoritative; for if we both refute the objections and leave the common opinions undisturbed, we shall have proved the case sufficiently.[8]

It may simply be coincidental that Aristotle offers this counsel at the outset of his discussion of wrongdoing and the manner in which it is to be explained. Yet we ought not miss the significance of the juxtaposition. If, as we have suggested, wrongdoing is the practical condition that makes theory matter, we ought to expect such theories that we have on offer to be mindful of the concerns of commonsense, not dismissive of them. We would be justifiably put-off by a moral theory that entailed that no one ever did wrong or that rendered the very notion incoherent. But this is precisely what we do find in the case of seemingly plausible moral theories. It is true, of course, that we might accept a theory of morality that was so revisionary that its practical significance consisted not in helping us to do right and avoid wrong but to disabuse us of a confused pre-theoretic understanding of such matters. It is no doubt possible that moral commonsense, with its conception of agents as frequently acting in ways subject to moral evaluation, is thoroughly confused. If so, then the varieties of morally-laden explanations of wrongdoing provided by commonsense—such as weakness, wickedness, and indifference, to name but a few—are likewise muddled and without extension. Perhaps this is what moral skepticism aims to show. Yet however high-minded the intentions of the moral skeptic, the vast majority of us cling to our moral practice, finding both comfort and intellectual satisfaction in the experience of significance that the moral outlook affords. Most of us take our commonsense talk of the moral dimension of existence quite seriously. Indeed, it seems to be what makes life *matter*. As such, the majority of those living the contemplative life aim not to debunk the moral life but to render it more precise and thereby more secure.

---

[8] *NE* VII 1 1145b1-7. See W. D. Ross, *Foundations of Ethics*, Oxford: Clarendon Press, 1939: 1-3, for discussion of this point.

It is all the more unfortunate, then, that theories appropriately bearing the label of *moral realism* strongly suggest that the commonsense picture is confused, in part if not in toto. This is so because realist theories, in various ways and to varying degrees, necessarily delimit the range of possible explanations of wrongdoing. As such, to varying degrees, they necessarily undermine their own practical significance. They become theories without a purpose. For it is but a short distance from the besmirching of explanations of some phenomenon to skepticism with regard to the existence of the phenomenon itself. By denying the plausibility of common accounts of wrongful behavior, realist theories suggest we doubt the coherence of the common notion of wrongdoing itself. But such skepticism is clearly not the intent of theorists who call themselves 'realists.' Regardless, realist-inspired skepticism should be found no more credible to the engaged moral practitioner than its more natural variant. If we are serious about morality then we will take wrongdoing to be the fundamental practical problem. It is therefore a theoretical challenge of the first order. The argument to be presented here is that it is a challenge that moral realism is ill prepared to meet.[9]

We need to set the terms of the discussion to follow at the outset. Our first priority is to clearly state the nature of our stalking horse and identify the proponents of the views that will here be critiqued. What, then, is moral realism? David McNaughton—who will figure prominently in the sequel—identifies the moral realist as one "who claims that moral values are to be found in the world, and that there are moral truths that we can discover."[10] More extensively, Richard Boyd identifies moral realism with the following three theses:

---

[9] My use of the term 'moral theory' has been rather loose. What I have in mind is more along the lines of a metaethical *approach* to morality—and its attendant moral psychology—rather than something recognizable as a normative theory.

[10] *Moral Vision* (Oxford: Blackwell Publishers, 1988): 1.

1. Moral statements are the sorts of statements which are (or which express propositions which are) true or false (or approximately true, largely false, etc.);

2. The truth or falsity (approximate truth...) of moral statements is largely independent of our moral opinions, theories, etc.;

3. Ordinary canons of moral reasoning—together with ordinary canons of scientific and everyday factual reasoning—constitute, under many circumstances at least, a reliable method for obtaining and improving (approximate) moral knowledge.[11]

What is most significant about moral realism, for present purposes at any rate, is that demands a distinctive construal of the *psychology* of the moral life, particularly of the nature of the psychological processes that underlie, or culminate in, the formation of a moral judgment (such as 'Jane is wicked,' or 'I ought to keep my word'). In the terminology of our day, that psychology is called 'cognitivist' (a label more steeped in history is 'rationalist'). Cognitivism claims that those psychological processes involved in making moral evaluations are essentially *descriptive* (hence cognitive). The essence of moral judgment, on this view, is the *representation* of certain things or phenomena (actions, motives, intentions, people, states of affairs) as *having* moral value. Moral judgments are essentially representational, on this view, because they are said to have "propositional or cognitive content."[12] Such representations—as are the locutions that convey them—are truth-evaluable (either one has correctly or accurately represented something as having a particular moral value, thereby rendering the representation true, or one has failed to do so, making one's representation false). Hence a leading proponent of *moral cognitivism* can articulate his view in terms quite similar to the moral realist. Thus David Wiggins says that moral cognitivism is the thesis

> that the judgments of morals are irreducibly cognitive in their aspiration. This is to say that moral judgments purport to represent moral knowledge (in the general and

[11] "How to be a Moral Realist," *Essays on Moral Realism*, edited by Geoffrey Sayre-McCord (Ithaca: Cornell University Press, 1988). Reprinted in *Moral Discourse and Practice*, edited by Stephen Darwall, Allan Gibbard, and Peter Railton, (New York and Oxford: Oxford University Press, 1997): 105.

[12] Stephen Darwall, *Philosophical Ethics* (Boulder: Westview Press, 1998): 71.

> ordinary sense of 'know') and that there is no other way for them to be seen by their authors, *qua* moral judgments, than as aimed at truth.[13]

The theses of moral realism and moral cognitivism are liable to be run together though they are nevertheless distinct.[14] One is a metaphysical thesis whereas the other is psychological. However, they are quite intimately related. A realist metaphysic would appear to demand a cognitivist psychology, and vice versa. In any event, it is with this psychological picture that I will primarily take issue, though I will return to the thesis of realism in Chapter four. Whatever might be said in favor of moral cognitivism (hereafter simply 'cognitivism'),[15] it simply does not square with our commonsense understanding of wrongdoing, nor does it permit of the variety of explanations of wrongdoing that commonsense affords. This renders such a psychology problematic, at best. Showing how and why this is so is the primary object of what follows.

In order to show the difficulties a cognitivist moral psychology faces as it attempts to explain wrongdoing, I will consider the views of a handful of self-proclaimed cognitivists

---

[13] "Moral Cognitivism, Moral Relativism, and Motivating Moral Beliefs," *Proceedings of the Aristotelian Society* (1991): 62. There are a number of different versions and names for theses in competition with cognitivism. Since all of them equally deny the truth of the cognitivist thesis they are frequently gathered together under the heading of *non-cognitivism*. This is unfortunate insofar as the general view is significant enough and of respectable pedigree to deserve a name of its own. Various versions of non-cognitivism go by the names 'emotivism,' 'prescriptivism,' 'expressivism,' 'quasi-realism,' 'conativism,' and, of course, classically, 'sentimentalism.' Both for reasons of convenience and for the fact that non-cognitivists generally agree that it is *conative/affective* states, such as desires, emotions, attitudes, and the like that are the essential basis of moral judgment, I will use the term 'conativism' as the alternative to the cognitivist position. Furthermore, we may add that since the states that the conativist take to provide the psychological basis of moral judgments are seen (by the conativist, anyway) to be non-representational, neither the states themselves nor the locutions used to express them are claimed to be truth-evaluable. For the conativist, moral locutions are a means of expression of the conative states underlying moral judgment.

[14] See David Wiggins, "Postscript" to *Needs, Values, Truth* (Oxford: Basil Blackwell, 1987): 329-1 for discussion of this point, as well as J.L Mackie's *Ethics: Inventing Right and Wrong* (London: Penguin Books, 1977). Mackie's "error theory" combines the truth of cognitivism about moral judgment with the falsity of realism about moral properties.

[15] Perhaps the strongest argument in favor of a cognitivist moral psychology is that it is consistent with certain features of the phenomenology of moral practice, particularly that fact that our moral judgments have (at the least) a surface structure which is fact-stating and property-ascribing. Furthermore, our moral discourse is steeped in the culture of truth and falsity. For instance, participants in the abortion debate take their opponents position to be *wrong*. They are making more than a moral claim when they do so.

who have discussed wrongdoing and the manner in which it is to be explained. These philosophers include John McDowell, David McNaughton, Jonathan Dancy, Michael Smith, David Brink, and Ronald Milo. All are cognitivists in that they take the nature of moral judgment to be representational and, thereby, truth-evaluable, but there are significant differences among their positions as well. The most important difference, for our purposes, concerns their views of the relation between moral judgment and motivation. Some (McDowell, McNaughton, Dancy, and Smith) argue that the relationship between moral judgment and motivation holds of conceptual necessity while others (Brink, Milo) maintain that it is only contingent.[16] The former, known as 'motivational internalists,' have argued that a necessary connection between moral judgment and motivation is needed to account for the essential practicality of morality, or, in other words, its motivational force.[17] According to this view, if an agent judges that some action is morally right then she is necessarily motivated to perform that action.[18] Likewise, when an agent judges some action to be wrong she necessarily has an aversion to performing it. The problem with positing such a strong connection, say the motivational externalists—those who hold that the relationship, insofar as there is one, between moral judgments and motivation is only contingent—is that it makes certain candidate explanations of wrongdoing conceptual

---

[16] An excellent starting point to the wealth of contemporary literature on this topic is W. K. Frankena's "Obligation and Motivation in Recent Moral Philosophy," in A. I Melden, ed., *Essays in Moral Philosophy* (Seattle: University of Washington Press, 1958): 40-81. Another informative discussion is found in Stephen Darwall's "Reasons, Motives, and the Demands of Morality: An Introduction," in *Moral Discourse and Practice*, Darwall, Gibbard, and Railton, 305-312. It is important to note that there are a number of related but different theses that go by the name 'internalism.' I am only concerned with that described in the text—motivational internalism (sometimes 'judgment' internalism)—the view that there is a necessary (conceptual) connection between moral *judgments* and motivation. I am not concerned with what is known as 'existence' internalism, the view that there is a necessary connection between moral/evaluative *properties* and motivation or simply having a reason to act.

[17] Recall an alternative way of understanding this notion on pg. 3.

[18] Which is not to say that she *will* perform it. There may be external obstacles to the action's performance as well as internal ones, such as contrary motivation.

impossibilities. Indeed, if moral judgment and motivation are necessarily connected, then any explanation of wrongdoing that relied on the absence of motivation to do what one judged right or the absence of an aversion to do what one judged wrong would be impossible.

The debate over the relation between moral judgment and motivation is more fine-grained than the internalist-externalist distinction would suggest. There is considerable friction among internalists themselves about how the judgment/motivation relation is to be construed and what underwrites it. We will, therefore, have reason to distinguish between *strong* internalists (McDowell, McNaughton, and Dancy) and *weak* internalists (Smith). The former hold the connection to be more intimate than the latter. This is because they understand moral judgments to be themselves motivational, whereas the latter sees them as distinct, though necessarily connected. It will be fruitful to consider these different cognitivist options (including the externalist option) because they reveal different problems that the cognitivist may face. Strong cognitive internalism will be the focus of the second chapter, while the weak version will be the target of Chapter three. The final chapter will consider cognitivism it its unalloyed form (externalism), and realism more generally. By that point I hope to have convinced the reader that the problems faced by cognitivist internalists, though magnified by their internalism, are not the result of internalism (*pace* the externalists) but of the cognitivist/realist framework itself.

A final preliminary point. It might well be asked if much sense can be made of *explaining* wrongdoing without previous adequate determination of the conditions of doing wrong. For it is notoriously possible to respond to the query of why one did wrong with the rejection of the question altogether. One person's immoral act is another's holy gesture. That different people often hold incompatible moral judgments about the same object

cannot be disputed. Whether at least one of those people *must* hold a judgment that is false can be questioned. However, this is an issue that we will leave largely to the periphery in what follows. The question of *whether* someone has done wrong is logically prior to the question framing the present discussion. *Given* that you and I take *X* to have done wrong, *how* are we to explain *X's* behavior? Answering this question in an intuitively satisfactory way poses a considerable challenge to moral philosophy as it is currently conceived.[19] As I have suggested, answering this question is something any moral theory worth having must do.

Explanations in Moral Practice

Take any action allowed to be wrong: willful murder, for instance. Examine it in all its lights, and see if you can find why someone might do it. A number of possibilities spring to mind. The murderer might be a jealous lover, a religious or political zealot, an impatient beneficiary of an inheritance or insurance policy, frightened of the possibility of an embarrassing revelation, or simply the sworn enemy of the victim. There are undoubtedly many more. All of these possibilities have been realized countless times over, as even the most casual observer of news reports can attest. But what are we, as observers of the affairs of people, doing when we offer such explanations? What are we saying about people when we claim that such explanations apply to them? At first glance it might appear that we are not saying much that anyone would find difficult to accept: we are claiming that people are capable of experiencing a range of feelings, emotions, desires, attitudes—in a word, motivations—that can bring them to intentionally end the life of another. Commonsense and philosophy both

[19] Though my discussion will be limited to cognitivist theories, it should be noted that all those discussed here are equally confident that no conativist (non-cognitivist) account can account for wrongdoing either. I will *very* briefly broach this at the end of Chapter four.

can have no quarrel with this. What we are speaking of here is nothing outside the realm of the human condition. We are capable of being motivated in countless ways to perform countless actions, even an action like murder.

Given what we have just said, it is understandable if one were to wonder why anyone would have thought that there is something problematic about explaining wrongdoing. People commit murders, rob and steal, lie and cheat, ignore those in need, abuse physically and emotionally others for reasons not too dissimilar from those for which they help others, speak the truth, give people what they are owed, and tend to those in need. Fear, anger, instrumental reasoning, and the like can as easily be implicated in acts of beneficence as well as maleficence. People can be shamed and embarrassed into the best of deeds. So why should we think there is a philosophically interesting issue here?

If we approach the issue of explaining wrongdoing in the manner just canvassed then it is likely that we will find little to trouble us. Explaining behavior—though no trivial task in itself—is *not* the primary concern of philosophers who take up the issue of wrongdoing. As we have so far discussed it, the issue is of explaining the behavior of an agent, behavior that *we* take to be wrong, and not necessarily behavior that the agent herself takes to be wrong.[20] Indeed, nothing has been said so far that would indicate *any* moral views on the part of the agent. But what happens when we do take such views into consideration? What influence, if any, do the moral views of the agent have on others' explanations of her (in our view) bad behavior? According to some philosophers, the agent's moral views place a substantive constraint on the form that an explanation of the agent's behavior can take. Similarly, philosophers may claim that what appears to be the appropriate explanation of the agent's

[20] The use of 'takes' is deliberate. I do not wish at present to distinguish between moral judgments in the narrow sense that implies something akin to belief and a more liberal sense of such judgments that would include conative states such as desire, emotion, and attitude. I will use 'takes' for this purpose throughout.

behavior determines the nature of the moral views to be attributed to her. The idea, expressed in the doctrine of motivational internalism (hereafter 'internalism') is that one's moral views and motivations are so intimately related that it would be unlikely, if not impossible, for an agent to *willfully* commit murder if she took murder to be morally wrong. And if it is indeed true that she willfully did murder, then she invariably took that particular action to be, at all events, morally permissible. Insofar as fear, anger, jealousy, and the like figured in the production of the behavior, this view claims that these motivations are not to be understood as operating independently of the agent's moral views. A goal here is to convince the reader—if she needs convincing—that such a view is implausible. No such constraints on the operation of our motivations should be accepted for to do so is to constrain both our explanations of wrongdoing and the content of the agent's moral views. Moreover, I aim to show what underlies the view that our moral explanations are so constrained. In so doing, I hope to make it clear that *any* philosopher who holds these underlying theses is thereby committed to holding an implausible view about explaining wrongdoing.

Those who argue there to be an intimate connection between moral views and motivation, i.e. internalists, are likely to offer a particular kind of explanation for wrongdoing. This is the explanation that Jean Hampton referred to as the "ignorance explanation."[21] According to this account, "immorality is due to ignorance of right and wrong."[22] This view was famously put forth by Socrates in *Protagoras* and the succeeding two sections will be devoted to understanding both how the ignorance thesis should be received

---

[21] Jean Hampton, "The Nature of Immorality," *Social Philosophy & Policy* 7 (1989). Reprinted in Amélie Oksenberg Rorty, *The Many Faces of Evil: Historical Perspectives* (London and New York: Routledge, 2001): 319-26. Quotation from pg. 319.

[22] Ibid., pg. 319.

and, by an analysis of Socrates' argument for it, why someone might find it plausible. Coming to grips with what gives the ignorance thesis its appeal is essential for what follows. The contemporary philosophers we will consider in subsequent chapters hold some or all of the theses held by Socrates. For this reason, I will argue, their attempts to explain wrongdoing are no better than his. If we ought to reject the Socratic explanation of wrongdoing as due to ignorance than we ought to reject contemporary theorists who offer it as well. Furthermore, understanding what underlies the ignorance thesis will also help us to understand the strange nature of Aristotle's account of wrongdoing and its relation to the Socratic account. In speaking of Socrates' view, Aristotle said, bluntly, that it "plainly contradicts the observed facts."[23] Despite such criticism, Aristotle went on to offer his own account the nature of which led him to the conclusion that "the position that Socrates sought to establish actually seems to result."[24] The problem for Socrates and Aristotle, as well as for contemporary cognitivists (internalists and externalists) is their moral psychology. Elizabeth Anscombe famously said that work in moral philosophy should cease "until we have an adequate philosophy of psychology, in which we are conspicuously lacking."[25] I agree with her on both counts: the philosophic enterprise of explaining moral practice has the philosophy of psychology as its foundation. Furthermore, the psychology favored by the dominant tradition in moral philosophy is, I will argue, out of step with the psychology underlying commonsense explanations of wrongdoing.

---

[23] *NE* VII 2 1145a 28-9.

[24] *NE* VII 3 1147b 14-15.

[25] G. E. M. Anscombe, "Modern Moral Philosophy," *Philosophy* 33 (1958). Reprinted in *Virtue Ethics*, edited by Roger Crisp and Michael Slote (Oxford: Oxford University Press, 1997): 26-44, quotation found on pg. 26. Citations of this article will be from this edition.

What are our commonsense views about wrongdoing? What types of explanations are to be found in our pre-theoretic moral practice? Perhaps no one has done more to attempt to make such views plain (as well as attempting to show how they are to be understood philosophically) than the contemporary philosopher Ronald Milo. Milo, fully appreciating the commonsense observation that there are many ways to do wrong, that our pre-theoretic moral practice is replete with rather fine-grained distinctions concerning the actions, motivations, and beliefs attributable to wrongdoers, offers a "typology" of immorality in his book *Immorality*. There he identifies no less than six different types of immoral behavior, some of which having distinct sub-species of their own. The main types to be found in our moral practice, according to Milo, include two different types of wickedness (perverse and preferential), amorality, moral negligence, moral weakness, and moral indifference. A brief comment on each will suffice for our purposes.

Wickedness may be defined as the deliberate doing of what is judged to be wrong "without any compunction or scruple."[26] But notice that in defining wickedness in this way we have not indicated whether the *agent* judges the action she deliberately performs to be wrong. At most the definition claims that an action deliberately performed by some agent is wrong in the judgment of critics. It is possible, then, that an agent that deliberately performs an act that critics judge to be wrong may in fact judge her act to be right. In that case we have an example of what Milo calls *perverse* wickedness, perverse because the agent perversely believes an action that is wrong is in fact right.

But what if the agent herself shares the critic's judgment that her deliberately chosen act is wrong? In that case we have an example of *preferential* wickedness. Milo sees the preferentially wicked agent as one believes that what she does is morally wrong, but does it

---

[26] Ronald. D. Milo, *Immorality* (Princeton: Princeton University Press, 1984): 29.

nevertheless because she prefers "the realization of some end to the avoidance of wrongdoing."[27] At the extreme, we can conceive of the preferentially wicked as one that deliberately does what she believes to be morally wrong *because* it is morally wrong. In this we would have the opposite of the conscientious moral agent that finds its clearest expression in Kant: the agent who does what she ought because she ought—the performance of duty for duty's sake.[28] This extreme form of preferential wickedness provides some of the content, I would assume, to the commonsense notion of *evil.*

The commonsense conception of *amorality*, Milo says, is one that applies to agents for whom moral considerations play no role in practical deliberations and do not figure in the motivation to act.[29] The reason for this could be either that such agents *lack* any moral convictions or, though they have moral convictions, those convictions are motivationally inert. David Brink has discussed the second possibility under the heading of amoralism, defining the amoralist as "someone who recognizes certain considerations as moral considerations and yet remains unmoved by them and sees no reason to act on them."[30] Milo, however, reserves the term 'amoral' for those agents that engage in wrongdoing due to lack of moral convictions, identifying those who have such convictions yet are unmoved by them as 'morally indifferent.'[31] Whatever label we wish to use, amorality and/or cold indifference is a particularly disturbing phenomenon. There is something frightening about

---

[27] Ibid., 29.

[28] It should be noted that Milo, echoing, among others, W.D. Ross, suggests that this extreme form of preferential wickedness is perhaps no more than an "'ideal' form of wickedness that is realized only by Satan and not by mere human beings" (7). For Ross' discussion see his *The Right and the Good* (Oxford: Clarendon Press, 1930), 163. Indeed, Milo, like David McNaughton in his *Moral Vision* (140-144), suggests that this ideal does not even apply to Satan. I will return to this issue later.

[29] Ibid., 56-7.

[30] *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press, 1989): 27.

[31] *Immorality*: 57.

the agent that seems to simply *not care* whether her actions are harmful or not, perhaps even more disconcerting than the agent who *intentionally* hurts someone. There is something alien, almost *inhuman* about such indifference and lack of concern. For this reason I think that amorality (understood broadly) also provides content to the commonsense notion of evil. A critic may judge the indifferent agent to be evil as easily—perhaps more so—as the preferentially wicked one.[32]

Perhaps no type of wrongdoing—and no type of blameworthy character—has received as much philosophical attention as *moral weakness*. 'Moral weakness,' just as the other frequently used terms 'weakness of will' and 'incontinence' suggest, implies a lack of self-control on the part of the wrongdoing agent. In cases of moral weakness (hereafter simply 'weakness') the agent is said to have moral convictions, is motivated to act in accordance with them (that is, motivated to do what she thinks is right and motivated to avoid what she thinks is wrong). Further, she is said to *intend* to act in accordance with them, yet because of strong appetites, desires, or powerful emotions, performs an act contrary to her intent. This has been thought problematic in two very different ways. It is thought problematic from both the commonsense and the philosophical perspective, and this accounts for its enduring interest.

Weakness poses a *practical* problem for commonsense: How is weakness to be avoided? Through our deliberations, choices, and intentions we hope to act successfully in the world, to navigate and coordinate between the various goals that we each have as well as with the goals of others. If we at times fail to act as we have intended our goals are placed in jeopardy and most likely will be frustrated. Weakness is a sign of unreliability, and is a failing that can be acutely personal: weakness is frequently self-ascribed. It is disturbing to feel that

---

[32] Milo also suggests that amorality and indifference may be thought worse than wickedness (246-7).

we cannot count on ourselves to do what we say we will. Nor is it likely that, if we exhibit weakness, we will be counted on by others. Being labeled 'undependable' is a particularly painful accusation for many of us to bear. Insofar as people turn to philosophy for help in practical matters, they want to be told what leads them to be weak so that they may rectify the problem and become strong.

It is to the great misfortune of commonsense and philosophy alike that the plea for practical help on the part of weak folk has fallen on philosophical deaf ears. Philosophers have not tended to see in weakness a practical problem for successful living but a theoretical problem for philosophy. Philosophers have rarely (if ever) asked how to avoid weakness but instead occupied themselves with the question of how weakness is *possible*. More alarming still, the frequent answer to that question is that it is not. Given the theses that some philosophers hold, either the idea of acting contrary to what one believes is right or contrary to how one intended or chose to act are conceptual confusions. In keeping with the methodological conservatism espoused in the previous section, I think it is more likely that the theses that prove incompatible with weakness are false rather than the commonsense conception of the phenomenon.

The final category of wrongdoing and immoral character that Milo considers is *moral negligence*. Following Aristotle, Milo sees the phenomenon of negligence as moral ignorance arising through the fault of the agent. Like in the case of weakness, the culprit is problematic appetites, desires, and emotions. But unlike weakness, where the effect of such states is the failure to act on one's avowed moral principles, in negligence the effect is the failure to judge that one's actions *fall under* those principles.[33] Though an agent knows that certain types of actions are, say, wrong, their lack of self-control leads them to fail to realize

[33] *Immorality*. 82.

that *this* action is one of that *type*. We will return to this kind of negligence when we discuss Aristotle's theory in the following section.

What does the richness of our moral practice tell us and what bearing does it have on moral theorizing? It will be convenient for our present efforts if we speak of three categories of wrongdoing—evil, weakness, and negligence—instead of six (or however many more). However, as we do we must keep in mind that such a limitation is not a *reduction* of the ways of wrongdoing but only a means of economy of expression (the discussion will be more fine-grained when necessary). Indeed, a substantive reduction in the ways of wrongdoing is something we should be on the alert to avoid. Insofar as moral practice employs these many different types of explanations of wrongdoing and denominates certain characters as blameworthy it is the role of moral theory to elucidate them and make them more precise and comprehensible, not to eliminate them. It is precisely this fault—that of eliminating pre-theoretically respectable ways of speaking about wrongdoing—that is the failure of the ignorance explanation. By claiming that all wrongdoing is properly explained as ignorance, the thesis falsifies much of our moral practice. A theory that reduces the ways of wrongdoing to one should be seen as problematic.

Let us take as fact the commonsense practice of distinguishing between evil, weakness, and negligence. What lies behind this practice? What is to be learned from it? The first thing to note is that there is a distinction in the *degree* of wrongdoing. Within the phenomena of wrongdoing itself, there is better and worse. For instance, we think of evil people as worse than those who are weak, and those who are weak are thought to be worse than those who are negligent. Our practice, then, involves a distinction in the degree of condemnation between evil, weakness, negligence, and any other form of wrongdoing or bad character that we condemn. To condemn an action or a person is to reject it in some fashion; it is to offer

a disavowal. But our disavowals, our condemnations, vary in strength and intensity and some are mitigated by a willingness to forgive, to sympathize, and to understand, while others remain pure and unalloyed repudiations. There are distinctions among ways of wrongdoing, and our condemnations of them are also moral in nature, indicating belief in their greater and lesser moral turpitude.

Our moral practice commits us to a further distinction, one that is *psychological* in nature and which underlies (which is not necessarily to say 'causes') the moral distinction. We tend to think of evil persons, those who we take to be the worst of all wrongdoers, as in control and sure of themselves, acting as they do because their aims, goals, and intentions demand it. Evil people, we think, perform acts of their own choosing. This contrasts rather sharply with our commonsense conception of those we consider weak. The weak seem like a house divided, engaged in internal strife and feeling as though they are on the losing side. The weak are *not* sure of themselves and do not always act in ways that they think are pursuant to their goals. At the very least we can say that the weak sometimes act in ways that they themselves do not wholly identify with. What is being distinguished here among these different kinds of wrongdoing is the psychological etiology of the resultant behavior. Moral weakness exhibits a conflicted etiology whereas evil (and negligence) does not. The morally weak are often said to be weak *of will*, suggesting that they act, in some sense, against their will, or unwillingly.[34] The evil (and the negligent) do as they do *willingly*; their actions are the ones they have chosen.

---

[34] I say that our moral practice has the morally weak acting against their will *in some sense*. Just what sense there is to acting against one's will is a large part of what I am calling the *philosophical* problem of weakness of will. I am at this point only stating what is the common sense or 'ordinary language' understanding of the phenomenon.

A further aspect of the psychological distinctions among ways of wrongdoing concerns what *follows* the action. For instance, the morally weak notoriously *regret* their behavior, as do the negligent when they realize the true nature of their act. The story is much different for the evil for whom, deathbed conversions aside, remorse is lacking.[35] We can represent these differences in the ways of wrongdoing that are implicit in our practice of both the doing and judging of wrong by the following schematic table.

	Negligence	Weakness	Evil
Action done willingly?	Yes	No	Yes
Regret?	Yes	Yes	No

A burden of ethical theory is to explain this practice of action and evaluation. How is it that we come to do the things we do? How is it that we come to make the judgments that we make? How could it be that we make such distinctions and how could it be that they are true? Any theory purporting to be *about* our practice, if it is to be adequate, must explain how we as agents can *engage* in these various ways of wrongdoing and indicate the ground on which we as critics make our judgments.

Socrates and the Ignorance Thesis

Let's now turn to the ignorance thesis of Socrates. That the Socrates of *Protagoras* held that wrongdoing was the result of ignorance is plain, but it is not obvious that he maintained this

[35] We might be tempted to think that this goes too far and that evil people can also suffer from doubt, uncertainty, and, perhaps, regret. We need to remember, however, that the categorization in question is a distillation of our practices of evaluation: evil is our category for the worst moral agents and the worst moral

view or that Plato did.[36] This, however, is not our concern here. Indeed, whether Socrates, Plato, or Aristotle actually held the ignorance thesis of wrongdoing or if their view is more nuanced or altogether different is immaterial for our purposes. Our aim is to elucidate a thesis about wrongdoing that is both false and—even if unwittingly—pervasive. The theory is the topic here, not its adherents. So, with the caveat that it is possible that neither Socrates nor Aristotle held the theory to be discussed, we will, for convenience, speak as if they did.

The chasm that exists between the theoretical perspective that spawns the ignorance thesis of wrongdoing and moral commonsense is starkly on display in the following passage from *Protagoras*:

> Come now, Protagoras, and reveal this about your mind: What do you think about knowledge? Do you go along with the majority or not? Most people think this way about it, that it is not a powerful thing, neither a leader nor a ruler. They do not think of it in that way at all; but rather in this way: while knowledge is often present in a man, what rules him is not knowledge but rather anything else—sometimes desire, sometimes pleasure, sometimes pain, at other times love, often fear; they think of his knowledge as being utterly dragged around by all these other things as if it were a slave. Now, does the matter seem like that to you, or does it seem to you that knowledge is a fine thing capable of ruling a person, and if someone were to know what is good or bad, then he would not be forced by anything to act otherwise than knowledge dictates, and intelligence would be sufficient to save a person?[37]

Given that commonsense holds that an agent may have sufficiently strong desires, emotions, passions, or other kinds of motivations that would enable her to act in ways other than as she believes (or knows) to be right or good, commonsense allows for the possibility that she

---

agents do *not* grapple with uncertainty or feel regret. At the very least, anyway, we should agree that those who do doubt and regret are not as evil as those who don't.

[36] An excellent discussion of the development of Plato's thoughts on the matter, showing that the view presented in *Protagoras* was by no means Plato's last word on the subject, can be found in J. J. Walsh, *Aristotle's Conception of Moral Weakness* (New York & London: Columbia University Press, 1960).

[37] Plato, *Protagoras*,, trans. by Stanley Lombardo and Karen Bell (Indanopolis: Hackett Publishing, 1992): 352b-c. Subsequent reference will appear parenthetically in the text.

be described as being morally weak, bad, wicked, indifferent, or evil (as well as being good and virtuous), or in a host of other ways. Commonsense can make such judgments only insofar as it does two things. The first is to makes a sharp distinction between "knowledge" (or belief), "intelligence," and such things as "desire," "pleasure," "pain," "love," and "fear." The second is to deny that members of the first group have so much power or influence in the mind so as to preclude the second group from ever being able to produce actions contrary to what "knowledge dictates." These are theses about psychological states and their operations. They are implicit elements of the commonsense picture of human nature in general and moral practice in particular. Only if there is a division of psychic labor where motivation is free of the draconian rule of the intellect could agents either fail, refrain, or reject doing what they understand or judge to be right, good, or best. The commonsense moral categories of weakness, wickedness, indifference, and the like only make sense on the presupposition of the psyche as an 'open society,' one where it is possible that any attitude, or motivational orientation, could be taken towards any belief.

For Socrates the psyche is not a free and open society but rather an intellectual dictatorship. Socrates insists that knowledge and intellect "rule" the psyche and "dictate" behavior. Due to this, most of the commonsense categories of wrongdoing are made impossible to satisfy. If it is the case that "if someone were to know what is good or bad, then he would not be forced by anything to act otherwise than knowledge dictates," then the only way to explain an agent's wrongdoing is in terms of her failure to know what is good or bad: wrongdoing is the result of ignorance of moral value (357d; 358c). That the etiology of wrongdoing cannot make exclusive reference to the motivations of an agent Socrates makes clear when he says

> Now, no one goes willingly toward the bad or what he believes to be bad; neither is it in human nature, so it seems, to want to go toward what one believes to be bad

> instead of to the good. And when he is forced to choose between one of two bad things, no one will choose the greater if he is able to choose the lesser (358c-d).

This is a claim about human psychological capacities. It holds that value judgments—evaluative beliefs—necessarily determine motivation. Socrates is, therefore, a (cognitive) internalist. The particular version of this thesis employed by Socrates in his argument is psychological hedonism.[38] However, the hedonism is not essential. As long as agents necessarily seek to act in ways that they represent as being good or best, whether they understand such values in terms of pleasure or not, they will not be capable of being motivated to do what they think is wrong. Without this ability, nothing but ignorance can serve as a plausible explanation of wrongdoing.

To claim that all wrongdoing is the result of ignorance may seem rather extraordinary. Indeed, I think that it *is* rather extraordinary, given what we observe daily with respect to human affairs and interaction, as well as the vicissitudes of our own concerns and motivations. However, we do not advance philosophically by dismissing out of hand claims that we find *prima facie* surprising—even implausible—without first considering both why those claims would be put forth and why they "have to be" false. If the ignorance thesis of wrongdoing is not only false but radically revisionist of commonsense, we need to ask why someone like Socrates would adhere to it. Fortunately, *Protagoras* holds a possible answer.

Immediately after Socrates asserts the version of internalist thesis quoted above, he turns to a member of his audience and offers the following.

> Well, then, is there something you call dread or fear? And I address this to you, Prodicus. I say that whether you call it fear or dread, it is an expectation of something bad (358d).

---

[38] See Gerasimos Xenophon Santas' "An Argument against Explanations of Weakness," *Philosophical Review* (1966), and reprinted in his *Socrates: Philosophy in Plato's Early Dialogues* (London: Routledge & Kegan Paul, 1979): ch. VII, esp. 206-7.

Whether we find this statement about fear (or dread—I take it that Socrates sees them as synonymous) innocuous or not, we oughtn't fail to understand its importance. What seems to be offered by Socrates is a highly controversial analysis of the psychological state of fear. What is being claimed is that fear—a state with motivational force, a conative state—is actually an evaluative belief. It is not that fear *accompanies*, either always or often, the expectation of something bad, nor that it is a *response* to such an expectation, but that it just *is* such an expectation.

The significance of analyzing fear as an expectation of something bad is great for the purposes of Socrates' argument against the common understanding of wrongdoing. Fear, it will be agreed, is motivational: fear generally moves the person feeling it to avoid the object of, or presumed cause of, that feeling. But if fear is properly understood as a negative evaluative belief then what is actually motivational is the evaluative belief. But now notice that if fear is naturally associated with avoidance then, on the present analysis, the evaluation of something as bad is naturally associated with avoidance: people naturally avoid what they believe to be bad.[39] And this is just to say what has already been said, namely, that human nature is such that it does not "willingly" go towards the bad or what it believes to be bad.

We have seen that if human nature is such to always be motivated to do what is believed to be good, right, or best and never towards what is believed to be bad, wrong, or worse,

---

[39] If we go further and presume that there is nothing so peculiar about fear that it should receive such an analysis but that other feelings, emotions, and passions should not, then we are confronted with a thesis about the nature of motivational states themselves: they are all evaluative beliefs. Whether Socrates held that *all* desiderative, or motivational, states are 'rational' in the way fear is said to be is a matter of debate. For discussion of this point see Thomas C. Birches and Nicholas D. Smith, *The Philosophy of Socrates*, Boulder: Westview Press, 2000, 179-181, and the citations therein. Whether or not Socrates voices a less inclusive view in other, later, dialogues, it does not seem to be present in *Protagoras*. On this point see Michael Frede's introduction to the Lombardo and Bell translation, *op. Cit.* xxix-xxx. Again, my concern is not with what Socrates actually held, especially at a time other than depicted in *Protagoras*, but with a view that is discussed in that dialogue. *If* one takes all motivational states to be evaluative beliefs, *then* one must claim that a person is necessarily motivated to act in accordance with those beliefs. If this is true, then all wrongdoing is the result of ignorance.

then ignorance becomes the only plausible explanation for wrongdoing. We now see that such a theory of human motivation naturally follows—trivially, in fact—from the thesis that motivational states such as desires, emotions, feelings, and the like are actually evaluative beliefs. If evaluative beliefs are essentially motivational (as desires, emotions, feelings, and the like are) then people's motivation is determined by their beliefs. With this 'identity thesis' of psychological states in place, the Socratic position that knowledge and intellect "dictates" behavior inevitably follows. It is worth dwelling on this point further.

I am suggesting that assuming that motivational and desiderative states are evaluative beliefs necessarily leads to the position that no one does wrong, or what they believe to be wrong, willingly, which, in turn, entails the ignorance thesis. So, I am suggesting that in holding the 'identity thesis' one must, if one is to be consistent, hold the ignorance thesis as well. Given that the ignorance thesis should be rejected on the grounds that it is incompatible and, ultimately, overly revisionary of commonsense, I am arguing that the identity thesis, which entails the ignorance thesis, is to be rejected as well.

The commonsense categories of wrongdoing, we have claimed, are predicated on the possibility of distinguishing between states such as beliefs—be they evaluative or not—and states such as desires, emotions, and appetites. After having distinguished such states, there is the further question of how they may be related to each other in any given instance. Different possibilities suggest themselves, some of which form the basis for various ways of wrongdoing we have mentioned. For instance, evil agents (be they preferentially wicked or amoral) are described as having desires or appetites uncompromisingly directed towards those objects or actions that are believed to be bad or wrong.[40] This contrasts with the weak

[40] The difference between the two being that the preferentially wicked agent's desires are directed at the object in virtue of the wrongful or bad characteristics it is believed to have whereas the amoral agent is presumably directed towards it for other reasons, she being indifferent to its moral properties.

for whom attraction to actions believed to be wrong is accompanied by an aversion to such actions, as well as a desire to do other, incompatible actions, thought to be right. But this distinction between the weak and the evil, based as it is on the possible ways that desires and the like can be related to evaluative beliefs, is impossible when the identity thesis is assumed. If desires and emotions just are evaluative beliefs then there is only one way for them to be 'related.' The commonsense distinctions represented in our moral practice are obliterated upon the adoption of the identity thesis.

Adequately explaining our moral practice with respect to wrongdoing requires faithfulness to the distinctions explicit in that practice. Unfortunately, many philosophers employ descriptions of categories of wrongdoing that fail to discriminate between them. This is particularly apparent in the way that a number of philosophers describe weakness—in a way that conflates it with wickedness. There is reason, I think, to fear that Socrates is guilty of just this conflation in his discussion. Consider what Socrates says after introducing the question of the plausibility of the commonsense picture of knowledge being "dragged about" by desire and the like (quoted above at the beginning of this section). When Protagoras responds by agreeing with Socrates' view that "if someone were to know what is good or bad, then he would not be forced by anything to act otherwise than knowledge dictates," Socrates responds with the following:

> You realize that most people aren't going to be convinced by us. *They are going to say that most people are unwilling to do what is best, even though they know what it is and are able to do it.* And when I have asked them the reason for this, they say that those who act that way do so because they are overcome by pleasure or pain or are being ruled by one of the things I referred to just now (352d-e, emphasis added).

The explanation offered by the majority, of being "overcome," is one that Socrates intends to show is "demonstrably false"(353a). The argument of *Protagoras* is, in other words, aimed at the phenomenon of weakness: the commonsense picture of this category of wrongdoing

is false. It is not the case that weak agents are overcome by desire, or pleasure, or pain, but by ignorance. But we should notice that whether it is false or not, the explanation of wrongdoing that is objected to has been rendered superfluous by what was said before it. It is not necessary that I be "overcome" by a passion or desire of any kind in order to be *unwilling* to do what I believe is best or right; I only need to not want to do it. I may have no interest at all in doing what I believe is right or, perhaps, I repudiate what I believe is right, aiming instead to do what I think is wrong. In such a case I would be properly described as wicked or evil, not weak.

Perhaps this is just an inadvertent mistake on Socrates' (or Plato's) part, a bit of sloppy writing not worth fussing about. However, as we shall see in subsequent chapters, this is a carelessness that a number of contemporary philosophers are also guilty of.[41] It is something of concern if for no other reason then that it betrays a lack of fidelity to commonsense. If a philosophical analysis of wrongdoing cannot clearly capture the distinctions of our moral practice then it is unlikely to be satisfactory. What the cause of this conflation is on the part of contemporary writers is something we will discuss later. As for Socrates, I think that it is arguably the adherence to the identity thesis that renders the distinction between being unwilling to do what is right and being overcome by desire to do something not right of no significance. We can see the difficulties facing Socrates when we try to reconstruct wickedness and weakness on the presupposition of his identity thesis.

---

[41] Much of the current Hare-Anscombe-Davidson-inspired debate does not adequately (if at all) distinguish between weakness and wickedness. Weakness is frequently characterized as *intentionally* not doing what is believed to be best or right. This is wickedness, not weakness. Examples of this conflation can be found in Donald Davidson's "How is Weakness of the Will Possible," reprinted in his *Essays on Actions and Events* (Oxford, Clarendon Press, 1980): 21-42, (conflation on pp. 21-22); David Wiggins, "Weakness of Will, Commensurability, and the Objects of Deliberation and Desire," *Proceedings of the Aristotelian Society*, 79, 251-277, (conflation on pg. 251); David Pears, *Motivated Irrationality*, Oxford, Clarendon Press, 1984, (conflation on pg. 264); and Robert Dunn, *The Possibility of Weakness of Will*, Indianapolis and Cambridge, Hackett Publishing Company, 1987, (conflation on pg. 1). Again, I will take this up in chapter two.

Consider wickedness. If motivational states such as desire and fear are evaluative beliefs then we must, if we are to make any pretense at capturing some aspect of commonsense, say that states that are understood to be *positively* motivational, or motivational *towards* something—states of attraction, such as desire and love—are expectations of good and *negatively* motivational states—states of aversion, such as fear—are expectations of bad. For convenience, we can borrow the terminology introduced by Donald Davidson and speak simply of pro- and con-attitudes.[42] Now, if the wicked are those who have a pro-attitude towards what they believe to be wrong, then, according to the Socratic version of the identity thesis, they believe it is good, or right, to do what they believe to be wrong. But this hardly makes sense. Either we must say that the agent thinks that one and the same thing is both right and wrong or that she thinks it is right to think it is wrong. Neither of these alternatives supports the commonsense picture. Let me explain.

Imagine that our wicked agent is considering murder. Commonsense has it that she knows (or simply believes) murder to be wrong but is motivated to do it anyway—she has a pro-attitude towards the murderous act. Now it might be said, correctly, that an agent might think that murder could have both positive and negative aspects to it—both right making and wrong making features. However, it is clear that what Socrates has in mind when he speaks of evaluative judgments that determine action are what we might call "all things considered" judgments. That is, to evaluate something as right or wrong, good or bad, is to evaluate it *on the whole*.[43] A thing or an act may contain both good and bad aspects, but the final determination of its value will depend on which aspects "outweigh" the others. If an

---

[42] See his "Actions, Reasons, and Causes," *Journal of Philosophy*, Vol. LX, (1963): 685-700.

[43] For more extensive discussion of this point, see Santas, "An Argument against Explanations of Weakness," *op. Cit.*: 200-204.

agent judges murder to be wrong it is because the negative aspects of murder are of greater total significance than the positive ones. But given that evaluative judgments are all things considered, the idea that the wicked agent thinks that murder is both right and wrong is ruled out.

This leaves us with the alternative that the wicked agent is making two *different* evaluative judgments, one concerning an object or action and another concerning the evaluation of the object or action. Why think this? It is reasonable to assume that if having a pro- or con-attitude can be understood as an evaluative judgment, then an evaluative judgment can be understood—at least for the sake of argument—as a pro- or con-attitude. So, if our agent believes murder is wrong then she has a con-attitude towards murder: she does not want to kill. Now if her wickedness is a function of her wanting—her having a pro-attitude—towards what she judges to be wrong, how is this to be understood in a way other than that she wants to not want to kill?[44] But wanting to not want to kill is an attitude we would encourage, not one we would condemn as wicked. There seems no way to capture the intuitive understanding of wickedness when motivation is analyzed as evaluative judgment.

If wickedness is impossible to capture on the identity thesis then weakness can not hope to fare better. Wickedness would seem to be the simpler concept in that the desire to do what is wrong is straightforward and uncompromised. With weakness there is conflict. The weak *want* to do what they believe is right yet do what is wrong. Or so we commonly think. What makes them do wrong? The power of their desire for pleasure, or their fear of pain or hardship. But if their fear of pain is actually the expectation of something bad then their expectation of something bad led them away from what they believed to be good. But,

---

[44] The only alternative reading is that she wants to do what she does not want to do. Presumably this is no better off than her thinking murder is right and wrong and, presumably, would be treated in the same manner.

again, if their expectation of something bad was strong enough to "overcome" their expectation of good, then, on the whole, their belief was that the action itself was bad. Likewise, it simply does not make sense, according to the Socratic psychology, to say that an agent believed some action to be right, wanted to do that action, believed some alternative open to her was not right, wanted to do that action as well, and gave in to the desire to do the worse action. For when this account is restated to accord with the terms of the identity thesis we get the following: an agent believed some action was right (and obviously was motivated to do it), believed some other course was right, and performed the latter. Why the latter? Because she expected greater good from it than the former. To say that she desired the pleasure she thought would come from the morally bad act is nonsensical when desiring something is the same as believing it to be good.

Commonsense distinguishes between weakness and wickedness but, given Socrates' views about the nature of psychological states, both are instances of the same problem: ignorance of the 'true' moral value of the acts in question. There seems little open to Socrates on which to base such a distinction, if he had cared to draw it. One suggestion would be to appeal to the etiology of the judgment that leads to action. Socrates might say that there is some indecision prior to their final evaluative judgment on the part of the weak: they go back and forth on which act is right until they decide, albeit mistakenly. This, presumably, distinguishes them from the wicked who are free of doubt. But why can't the wicked be unsure at some previous point as well? Why think that they must never have wondered or doubted what they ought to do? Why couldn't the wicked, intent on doing what was right be (at some point, at least) unsure about what *is* right (and ultimately make an incorrect determination)? Yet even if this is allowed, as I think it should be, indecision prior to eventual belief formation is a poor substitute for conflicted motivation in the wake of

belief formation. The commonsense distinction between weakness and wickedness cannot be captured in this way.

We said that it was a part of our moral practice to distinguish different ways of wrongdoing by the presence or absence of regret. And this might suggest a second way for Socrates to ground a difference between ways of wrongdoing. However, regret, too, needs reconstruction on the assumption of the identity thesis. Agents invariably do what they think is right, or best. Frequently, though, they find that what they thought was best was not—their expectations of good were ill founded. When this occurs, it is normal for agents to regret their choice, to say that they always wanted to do what was right. Indeed they did, Socrates would say. They always wanted to do what was right, they did what they thought was right, and were all along mistaken about what was right. Ignorance, not wayward motivation, is the cause of their wrongdoing. And what of the wicked in this regard? They, as we have said, do not, normally, regret their behavior. How is *this* to be accounted for on the identity thesis? We would seem to have to say that wicked agents either remain ignorant about their choice, continuing to think that it was right, or come to learn of their mistake yet subsequently endorse it after the fact. Neither of these alternatives seems correct, however. The latter is impossible if 'endorsement' is understood to be a pro-attitude: one simply cannot have a pro-attitude toward (believe to be good) what she believes to be bad. The former are in danger of being dismissed as merely stupid, maintaining their "ignorance" of right and wrong even after the results are in. Neither, needless to say, is what our moral practice takes wicked agents to be. True, it is a datum of that practice that (at least many) wicked agents stick by their choices, and it is equally clear that many wicked agents claim that what they did was right. Moral practice, however, has a far richer account of this than the simple analysis claiming such agents to be ignorant of what is truly right and wrong. Far

from illuminating our moral practice, the ignorance explanation oversimplifies to the point of obfuscation. The Socratic account of wrongdoing presented in *Protagoras* ought to be rejected. The ignorance thesis is an impoverished account of why people do wrong.

The purpose of this discussion of the Socratic thesis has not, primarily, been to dissuade anyone of believing in its plausibility. It is fair to assume that few philosophers would embrace such an extreme account (indeed, Aristotle felt compelled to address the issue of its *prima facie* absurdity). The point, rather, was to expose the underlying psychological theses that lead one to hold such a view. By exposing the underpinnings of a thesis that runs so counter to commonsense we can more readily see how contemporary philosophers may, however unwittingly, find themselves similarly committed to problematic construals of our moral practice of explaining wrongdoing. Indeed, if the arguments presented here are correct, the difficulties facing cognitivism in dealing with the problem of wrongdoing are entirely conceptual in nature. The remaining chapters aim to show that these difficulties are inexorable. Before I turn to that, however, there is one last item to be mined from the Socratic argument, in fact the most significant of all. As we are about to see, the Socratic thesis that *all* wrongdoing is the result of ignorance is not simply out of step with commonsense (as Socrates plainly saw) but is actually incoherent. The real problem of wrongdoing is impossible to articulate with the moral psychology Socrates employs.

Ignorance, Negligence, and Motivation

I have been so far content to use the standard English translation of **ἀμαθία**, namely 'ignorance,' in discussing the Socratic view. However, this is actually quite problematic, for it cannot simply be 'ignorance' that Socrates has in mind in his discussion with Protagoras

unless it is his aim to argue for a thesis even more radical than is normally attributed to him. Unless we are to understand Socrates as arguing either (A) that there is no point to the institution of ascribing blame and meting out punishment because no one is *ever* culpable for their behavior (since ignorance is an excusing condition), or, (B) that though there is culpability, one is strictly liable for failure to do the objectively right act (there being no extension to the concept of a subjectively wrong one), we must understand Socrates to be arguing that all (blameworthy) wrongdoing is the result of *culpable* ignorance, that is to say negligence.[45]

As much as the various categories of wrongdoing are stabile fixtures of our pre-philosophical commonsense, so is the excusing nature of ignorance. Agents are not blamed or punished if they are *blamelessly* ignorant of the moral turpitude of their actions. To do otherwise would be to treat the agent unfairly. Nevertheless, agents may be blamed *for* their ignorance which is to say that they are not so much blamed for their having made a mistake but for their *being* ignorant. Moral negligence, as we saw from Milo's typology of commonsense, is the failure to know, do, or prevent that which would have averted the, as such, ignorantly performed wrongful act. Such ignorance is culpable insofar as the agent *should* have known the act was wrong *because they should have* known, done, or prevented that which would have either revealed the 'true' moral nature of the act or, at any rate, would have averted it.[46] It is helpful to think of negligence as a *relationship* between two 'acts.' As Holly Smith has put it, there is the initial act, a "benighting act," and the subsequent

---

[45] I assume that the distinction between subjective and objective moral value is sufficiently intuitive and clear for the purposes at hand however theoretically problematic—not to say superfluous—the concept of objective moral value may be. For an illuminating discussion of the distinction see W. D. Ross, *Foundations of Ethics* (Oxford: Clarendon Press, 1939), 146-67. I thank Professor Martin Harvey for pointing out the significance of this distinction, as well as strict liability, in this context.

[46] See H.L.A. Hart's "Negligence, *Mens Rea* and Criminal Responsibility," in *Punishment and Responsibility* (Oxford: Clarendon Press, 1968), 137.

"unwitting wrongful act."[47] If the benighting act (which may frequently be an omission) is not itself culpable, the resulting unwitting wrongful act will not be so either. *Some* act, whether performed in ignorance or not, must be blameworthy or no act will. Either Socrates thinks that *no one* engages in blameworthy behavior or they do so only from negligence.

Let us assume, then, that Socrates did not intend to deny the very existence of blameworthy behavior. But how, exactly, are we to distinguish (excusable) ignorance from negligence? There are, it seems, (at least) two ways to characterize this distinction. One is in terms of the *manner* or *source* of the ignorance, while the other is in terms of the object of ignorance. The former suggests that the distinction is rooted in the different answers to the question of *why* the agent is ignorant. The latter suggests that the distinction is rooted in different answers to the question of *what* the agent is ignorant about. Both ways of characterizing the difference converge, I think, in Aristotle's discussions of ignorance and negligence in *NE* III.

The early chapters (1-5) of *NE* III concern the concept of moral responsibility. Agents that are morally responsible are subject to praise and blame; they (along with their actions and passions) are proper objects of the evaluations of critics. The criterion of moral responsibility, for Aristotle, is that the "passion" or "action" being evaluated by the critic (and hence the agent feeling the passion or performing the action) be felt or performed "voluntarily."[48] Just what makes the feeling of a passion or the performance of an action voluntary is initially discussed in negative terms, through a discussion of those factors that

---

[47] "Culpable Ignorance," *The Philosophical Review*, Vol. XCII, No. 4 (1983), 543-571.

[48] *NE* III 1 1109b 30-35. For the remainder of the chapter, subsequent citations to the *Nicomachean Ethics* will be given parenthetically in the text.

serve to render passions and actions *in*voluntary, and therefore not subject to the praise and blame of critics. Two sources, or conditions, under which passions or actions are said to be involuntary are compulsion and ignorance. An agent is said to be under compulsion when "the moving principle is outside, being a principle in which nothing is contributed by the person who is acting or is feeling the passion" (1110a 2-3).

The ensuing discussion of ignorance introduces a number of distinctions. The first, interestingly, has nothing to do with ignorance *per se*, but with the attitude of the agent in the circumstances of the (potential) post-action dissipation of ignorance. An agent is said to have acted (or felt a passion) involuntarily if, after realizing the ignorance under which she acted (or felt), she feels "pain and repentance" (1110b 18-9), otherwise her action (or feeling) is said to be "*not* voluntary" (1110b 17). The importance of this distinction would seem to be intuitive: even if an agent 'did not know' that her action was harmful or wrong but, upon learning that it was, was not in any way disturbed, upset, or regretful, we would find her and her behavior objectionable and subject to censure. Though it would be true to say that she acted without knowledge, which is to say ignorantly, she does not subsequently *reject* or *deny* her action. Indeed, it is as if she were saying that *if* she had known that her action was of such a kind she still *would* have chosen it knowingly. Such an agent is criticized, not excused. To be excused it is necessary that an agent be repentant.

Repentance may be necessary for ignorance to be exculpatory but it is not sufficient. This is made clear by Aristotle by the other distinctions he makes in his discussion. The next distinction we should note concerns the *objects* of knowledge or, to put it the other way round, those things that an agent could be ignorant of. "Now every wicked man is ignorant of what he ought to do and what he ought to abstain from," said Aristotle, "and it is reason

of error of this kind that men become unjust and in general bad" (1110b 27-29). But Aristotle quickly goes on to add the following.

> [B]ut the term 'involuntary' tends not to be used not if a man is ignorant of what is to his advantage—for it is not mistaken purpose that causes involuntary action (it leads rather to wickedness), nor ignorance of the universal (for *that* men are *blamed*), but ignorance of particulars, i.e. of the circumstances of the action and the objects with which it is concerned. For it is on these that both pity and pardon depend, since the person who is ignorant of any of these acts involuntarily. (1110b 30-1111a 2. Emphasis in the translation.)

We have here a three-fold distinction among objects of knowledge, between purposes, or the ends of action, universals, and particulars. Ignorance of the first two (purposes and universals) leads a person to be wicked, bad, or unjust—that is, to be blamed—and is therefore inexcusable.[49] It is only ignorance of the particulars of action that Aristotle finds that commonsense is willing to excuse. That is, an agent may be excused for being ignorant of a particular (perhaps quite relevant) description of her action, whom it would effect, what she is using to perform it, what the consequences will likely be, and the manner in which she is doing it.[50]

I said that an agent "may" be excused for being ignorant of such particulars, for it is by no means necessary that she be so. Aristotle puts the point quite clearly when he says

> Indeed, we punish a man for his very ignorance, if he is thought responsible for his ignorance, as when penalties are doubled in the case of drunkenness; for the moving principle is in the man himself, since he had the power of not getting drunk and the getting drunk was the cause of his ignorance. And we punish those who are ignorant of anything in the laws that they ought to know and that is not difficult, and so too in the case of anything else that they are thought to be ignorant of through carelessness; we assume that it is in their power not to be ignorant, since they have the power of taking care (1113b 30-1114a 3).

---

[49] Exactly how "purposes," or what a person "ought to do and what he ought to abstain from," on the one hand, and "universals" are different for Aristotle is difficult to determine. Indeed, it is arguable that the universal premise of the practical syllogism is an expression of precisely what the purpose of the agent is, that is what he thinks he ought to pursue. I will return to this point later. Suffice it for now to say that whether these objects are distinct or not, ignorance of them is not excused.

[50] *NE* III 1 1111a 4-6.

What we have here is a distinction in the manner that the ignorance comes about, or, put in other terms, in the *cause* of the ignorance. An agent's ignorance is exculpatory when, and only when, the cause of her ignorance is 'outside,' or 'external' to her. If, however, *she herself* is the cause of her own ignorance (it being then 'internal'), she is not excused but blamed and subject to punishment.[51]

The picture we get from Aristotle's discussion is that ignorance is blameworthy—and therefore a case of negligence—when the ignorance is the agent's own fault. When someone else (through, say, dishonesty), or the external world itself (not all of the facts of our circumstances are knowable), brings about, or is the source of an agent's ignorance, the agent is excused; we think the agent could not have prevented the ignorance with which she acted. In this I think Aristotle has expressed a clear (and timeless) tenet of commonsense. But Aristotle goes further and limits the external influence on an agent's ability to know by reference to the objects of knowledge. To say that ignorance of purposes or of universal features of actions is *never* excusable is to say that such ignorance *never* has an external cause: the wicked, the unjust, and the bad are so because of *their* negligence. They are ignorant of what they ought to do and of what *kinds*, or *types*, of actions are good and bad, and their ignorance is—necessarily—due to themselves.[52]

---

[51] See Milo's *Immorality*, 31-2; Sarah Broadie, *Ethics With Aristotle* (New York & Oxford: Oxford University Press, 1991): 148.

[52] To speak of action kinds or types is to speak of the universal features of actions, and the relevant kinds for our discussion (as well as Aristotle's) concerns moral kinds, e.g. that lying is wrong. It is Aristotle's claim (which is to say that it is a part of our moral practice, according to him) that we are not willing to excuse a person that claims to be ignorant that lying or stealing is wrong, or that cowardice is a vice, but we *may* excuse a person who fails to know that *this particular act* is a lie or is cowardly. We will excuse such ignorance if we feel that the reason for the ignorance is external to the agent. We will return to this in the discussion of Aristotle's treatment of wrongdoing in next chapter.

Negligence is a cause of wrongdoing that may, if habituated, grow to be a deep-seated character flaw. The agent, through laziness, carelessness, lack of emotional control, disinterestedness, or some other personality trait fails to get morally relevant information. Without such information, doing the right thing is, at best, a matter of luck; more likely wrongdoing will result. If the Socratic account of wrongdoing is to be understood as an explanation of *immorality*, i.e. blameworthy wrongdoing, then it must be negligence that he has in mind. But now we see just how difficult the Socratic position is. The "mental dereliction"[53] that is the cause of the ignorance is, as we have seen, the result of some problematic conative/affective state. It is a *motivational* problem. But this would seem impossible to explain given the account of human motivation that Socrates employs. What underwrites the Socratic claim that wrongdoing is the result of ignorance (negligence), and *not* problematic motivations (as "most people" would have it), is his claim that "no one goes willingly towards the bad, or what he believes to be bad." Socrates, as we have seen, is an (motivational) internalist. But it is precisely this internalism that prevents him from explaining the possibility of negligence. Put baldly, negligence is ignorance that results from insufficient motivation to do or learn what one ought in order to avert some subsequent wrongful act. But this is to lay the moral fault at the feet of motivation, precisely what Socrates was arguing against.

The difficulty becomes manifest when we ask what is *wrong* with the negligent agent's motivation, for what answer can be given? If we assume that no one does wrong willingly then we can only say that the agent's motivations are problematic to the extent that they are directed at a mistaken conception of the good. But this way lies a vicious circle. Either the agent was insufficiently motivated to prevent subsequent wrongdoing because she simply did

---

[53] The phrase is from Holly Smith, "Culpable Ignorance," *Op. cit.* p. 547.

not *realize* what was at stake, in which case she would be excused for her ignorance, or she *did* realize the risk that would result from neglect but was *nonetheless* insufficiently motivated, and therefore guilty of negligence. But this second possibility is only open to us if we allow for the evaluation of motivations themselves. That is, we must allow that motivations are objects of evaluation (they can be praiseworthy or blameworthy) and are evaluable *independently* of the moral conception of the objects to which they are directed. To allow for this, however, is to deny the conceptual prohibition on willful wrongdoing. If we are to make sense of wrongdoing *at all*, then we cannot think of motivation as necessarily tied to moral judgment. Cognitive internalism is an unacceptable constraint on theorizing about our moral practice of explaining wrongdoing. Indeed, it falsifies the very phenomena it is meant to explain. Whether all wrongdoing is the result of negligence or not, cognitive internalism is false. But if negligence is possible, why not weakness, wickedness, and the like? If any of these were possible it would seem that they all would be so. What I am arguing for here is that these ways of wrongdoing are incompatible with cognitive internalism. If, as moral theorists, we hope to address a practical need for our services, then we *need* wrongdoers. Wrongdoers, though, need the freedom to do wrong, and this is precisely what Socrates' internalism denies them. At the root of blameworthy behavior is blameworthy motivation. The Socratic argument of *Protagoras* provides an instructive example of the cost of denying that truth. As we are about to see this is a lesson that has not always been heeded.

2

What Moral Judgments are Not

> The exposure of fallacious ethical arguments is, however, a task that which seems to be necessary to perform anew in every age. It is something like housekeeping, or lawn-mowing, or shaving...Even when we know beforehand that some system must be fallacious—that what it sets out to do, simply cannot be done—we learn something in the effort to discover just where the fallacy lies.
>
> —Arthur N. Prior, *Logic and the Basis of Ethics*[1]

We ended the previous chapter by arguing that the moral psychological thesis known as cognitive internalism is incompatible with our commonsense understanding of wrongdoing. Wrongful behavior, it was urged, is rooted in blameworthy motivation: whether we are speaking of weakness, wickedness, indifference, or negligence, the cause of such phenomena is to be found in the motivations of the wrongdoing agent. An agent's motivation can be blamed for being insufficient, inconstant, or simply nonexistent, not to mention malicious, spiteful, jealous, conniving, ungrateful, and more. But such criticisms of motivation are predicated on the independence of motivation from moral judgment. If moral judgment and motivation are necessarily connected so that an agent that judges some action to be morally right is necessarily motivated[2] to perform that action, then it is not clear what sense can be given to claims of the 'inconstancy' or 'nonexistence' or 'spitefulness' of motivation. None, at any rate, that carries a sense of moral disapprobation. Blameworthy behavior, I am arguing, rests on blameworthy motivation, and that entails the falsity of cognitive internalism.

---

[1] Arthur N. Prior, *Logic and the Basis of Ethics* (Oxford: Clarendon Press, 1949/1965), x-xi.

[2] Which is not to say necessarily moved.

In this chapter we shall begin to consider in detail why this is so. Our discussion will begin with a tour of the 'logical space' of moral psychology itself. As we shall see, some accounts of moral judgment cohere with accounts of the relation between such judgments and motivation. Other accounts of the two are quite at odds. Apart from questions concerning what we might call the 'internal consistency' of these theses, we need to discover whether and how these accounts cohere with commonsense views of the subject. We shall then see why the combination of cognitivism about moral judgment and internalism about motivation is at odds with our commonsense moral picture. Since that commonsense picture provides the material that philosophical theory is meant to address, this bodes ill for the cognitive internalist. As we have just seen with our discussion of Socrates, the truth of the cognitive internalist thesis entails the non-existence of the phenomena the thesis is supposed to explain. Nevertheless, cognitive internalism thrives in the contemporary philosophic landscape. The bulk of this chapter—and the next—will address the work of those philosophers who both adhere to some version of cognitive internalism and work on the problem of explaining wrongdoing. The burden of my argument will be to show that their attempts to explain wrongdoing won't prove successful.

Moral Psychology: A Brief Survey

Cognitivism, as we saw in the previous chapter, is a thesis about the nature of moral judgment. It is an account of what is going on in the formation of such judgments and what the aim of such judgments are. It is also an account that has consequences concerning the semantic status of both the mental states that comprise those judgments and the locutions used to express them. Moral judgments, according to cognitivism, involve the description and representation of the moral features of the world (whether those features are natural or

not), aim at the correct, or true, representation of those features, and are comprised of what we would normally call 'beliefs' which are, along with those statements expressing beliefs, truth-evaluable.

Cognitivists are divided on the relationship between moral judgments—as they understand them—and motivation. Depending on whether they are internalists or externalists, cognitivists offer differing accounts of the various categories of wrongdoing we identified in the previous chapter. Indeed, much of the internalist/externalist debate is fought over the proper understanding of the ways of wrongdoing. As has been stated, the primary thesis of this work is that neither version of cognitivism is well suited to provide plausible explanations and accounts of wrongdoing. This, though, is a claim that needs to be developed in stages. The first stage, begun in earnest in this chapter (and completed in the next) is to argue that if one is to be a cognitivist then one ought to be an externalist. Cognitivism and internalism do not mix. The commonsense moral realm, the perspective that provides the notion of moral wrongdoing with its natural setting, is incompatible with the picture that develops from the combination of cognitivism and internalism. Cognitivism or internalism is true, but not both.

When I say that either cognitivism or internalism is true, I mean to say that one or the other *must* be true. This becomes readily apparent when we consider the alternatives to cognitivism and internalism (conativism and externalism). Those who do not find cognitivism a viable thesis of the nature of moral judgment are likely to adhere to some version or other of *conativism*, the thesis that takes moral judgments to be a matter of one's conative/affective (more colloquially, *motivational*) orientation towards certain phenomena. Moral value, on such an account, is a projection of the moral practitioner's orientation, not a property of the evaluated things themselves. Standard moral locutions are understood as the

medium for expressing such orientations, and are, therefore, like the orientations they express, neither true nor false.

Consider the claim that killing innocents is wrong. According to cognitivism, this statement is one whose veracity is in question; it is either true or false that killing innocents is wrong. The statement either correctly or incorrectly depicts its subject as having a certain property (in this case the property of being wrong). The conativist, in contrast, denies that truth and falsity are aptly predicated of such a statement. Instead of attributing a particular moral property, that of being wrong, to the killing of innocents, the statement is expressing something about someone and *not, per se*, about the killing of innocents. That person may simply be the speaker herself, though the speaker may take herself to be speaking for others (perhaps an entire community or significant majority of its members). These differences in the semantic treatment of moral statements are derived from the differing views cognitivists and conativists have about the identity (really the nature) of those mental states such statements are employed to communicate.

It is commonplace in philosophical psychology to claim that creatures that exhibit *agency* must possess (at least) two seemingly different psychological capacities. They must be able to (A) represent to themselves their situation, any action that they might perform, and the subsequent alterations to their situation that performance of an action might bring, and (B) impel or induce their performance of some action. In the technical jargon of the day, they must have *cognitive* and *conative* capacities, respectively.[3] In somewhat less technical terms,

---

[3] Agreement is much less clear on whether agents also require an *affective* capacity, that is a capacity to 'feel.' Indeed, it is a controversial topic in the philosophy of mind whether such 'phenomenological' or 'qualitative' states actually exist and, if so, whether they are in the domain of psychology or physiology. (True, 'affective' has traditionally been used to speak of states like emotion, more so than states like (physical) pain and the experience of seeing green. However, it is common to give the identity conditions of all such states in terms of introspective features that indicate 'what it is like' to be in such a state.) I am inclined to think that, as a matter of fact, *human* agents, as a result of their physiology, have affective states. I also think, as I take did Hume, that affective states are best understood as a sub-category of conative states. Hence, some, but not all, motivational

agents must be able to have, compare, and contrast thoughts; form beliefs; and make implications; as well as have and prioritize cares, concerns, and goals, reactions to situations (both perceived and conceived); and, in general, be motivated. An agent is a creature capable of action and neither unmoved thinkers nor unthinking movers can act.[4]

The difference between cognitivism and conativism with reference to moral judgment is simply this difference between kinds of psychological capacities applied to the moral case. The cognitivist claims that moral statements like "killing innocents is wrong" communicate the thought, or belief, that killing innocents is wrong. Such a mental state is a cognitive, or representational, state and the semantic properties of the statement "killing innocents is wrong" are inherited from the semantic properties of the belief that it communicates (which semantic properties are, in turn, inherited from the semantic properties of the proposition to which the belief stands related). The conativist's denial of the truth-aptness of the statement "killing innocents is wrong" is due to the fact that, for her, that statement is communicating a conative, or motivational, state (such as disgust, anger, horror, aversion—or a disposition to have such states—towards such an act, or the thought of such an act), a state that is not itself truth-apt.

As we have said, this distinction between the cognitive and the conative, the representational and the motivational, is one that is common in philosophical discussions of

---

states have a 'feel,' and all states that have a 'feel' are motivational. I will not be arguing for either of these points here. As for theoretical similarities in attempting to understand qualia such as pains and affective states such as emotions, consider reference to Daniel Dennett's "Why You Can't Make a Computer that Feels Pain," in *Brainstorms* (Cambridge: Bradford Books, 1978), 190-229, found in Paul Griffiths' discussion of emotions in his *What Emotions Really Are* (Chicago: University of Chicago Press, 1997), 226, 256. As for an interpretation of Hume that takes him to identify conative states ('passions') as motivational states that may, or may not, have phenomenological features, see Michael Smith, "The Humean Theory of Motivation," *Mind*, 1987, 36-61, and his revised presentation in *The Moral Problem* (Malden and Oxford: Blackwell Publishers, 1994), ch. 4.

[4] Of course 'action' here is being understood in a very narrowly defined sense (physical behavior with mental causes).

psychology from Socrates to the present. One currently popular way of metaphorically capturing this distinction is attributed to Elizabeth Anscombe, involving what she termed 'direction of fit.'[5] The idea is captured nicely by Mark Platts in the defense of cognitivism he offers in his *Ways of Meaning*.[6] There he states the following.

> Miss Anscombe, in her work on intention, has drawn a broad distinction between two *kinds* of mental state, factual belief being the prime exemplar of one kind and desire a prime exemplar of the other...The distinction is in terms of the *direction of fit* of mental states and the world. Beliefs aim at the true, and their being true is their fitting the world; falsity is a decisive failing in a belief, and false beliefs should be discarded; beliefs should be changed to fit with the world, not vice versa. Desires aim at realisation, and their realisation is the world fitting with them; the fact that the indicative content of a desire is not realised in the world is not yet a failing *in the desire*, and not yet any reason to discard the desire. The world, crudely, should be changed to fit with our desires, not vice versa.[7]

This is a helpful, yet limited, way to distinguish the cognitive from the conative. First, as Platts is quick to emphasize, it is only a metaphor. It is important that we be able to cash out this metaphor in more concrete terms, terms that will enable us to indicate the *real* difference between cognition and conation, a difference that both cognitivists and conativists think *matters*.

A second worry, at least *prima facie*, is that the metaphor seems somewhat limited on the conativist's end. For the cognitivist it is clear that the 'prime exemplar' of the kind of state she has in mind that is the basis of moral judgment is a belief (even if cognitivists such as Socrates—and, indeed, McDowell—frequently speak of 'knowledge') and here the intuitiveness of mind-to-world direction of fit is quite strong; the aim of belief is to correctly represent reality. For the conativist, however, the pride of place that is here given to desire

---

[5] G. E. M. Anscombe, *Intention* (Oxford: Basil Blackwell, 1957).

[6] Mark Platts, *Ways of Meaning* (London: Routledge & Kegan Paul, 1979), 256-7.

[7] Ibid., 256-7.

is not so clearly warranted. Conativists have offered a variety of different candidate states as serving as the basis of moral judgments. For instance, Hume and Adam Smith speak of feelings and sentiments, Ayer of emotions, and Blackburn of concerns and attitudes.[8] The conativist might also wish to speak of drives, intentions, and moods. There are probably more states that might be more aptly identified as 'motivational' than as 'representational,' conative as opposed to cognitive. But the world-to-mind direction of fit that desires are supposed to be 'prime exemplars' of is much less intuitively apt to these other states.

Perhaps this is not such a problem for the conativist. She might say that emotions, concerns, attitudes, and the like are related to desires. Indeed, when an agent is angry or frightened, the 'natural' inclination is to alter or address one's situation to assuage one's feelings, not attempt to alter one's emotional makeup to more harmoniously approach the world.[9] Likewise, one's concerns and attitudes seem, in an important sense, independent of the current state of affairs. Concern for the welfare of one's children would seem to survive (and we might add *ought* to survive) the removal of any dangers, threats, or impediments. And our attitudes, if they are stable aspects of personality, continue in many instances despite changes in our situation. Yet even when we do speak (as we should) of attitudes changing as a result of changes in situation, there is not the inclination to speak of attitudes (or concerns, emotions, desires, etc.) as being *true*. This difference between beliefs and motivational states appears clear and indicative of something quite basic, perhaps what the metaphor of direction of fit was meant to capture: our beliefs *report* the world to us, tell us

---

[8] David Hume, *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd Ed revised by P. H. Nidditch (Oxford: Clarendon Press, 1888/1978); Adam Smith, *The Theory of Moral Sentiments*, edited by D. D. Raphael and A. L. Macfie (Indianapolis: Liberty Fund, 1984); A. J. Ayer, *Language, Truth, and Logic* (New York: Dover, 1952); Simon Blackburn, *Ruling Passions* (Oxford: Clarendon Press, 1998).

[9] Though this is precisely what one might do if one is convinced that altering the world in the appropriate way is either impossible or involves prohibitive cost.

how it *is*. Desires, emotions, attitudes, and such, on the other hand, determine how we *engage* the world: how we deal with it, react to it, and act in it.

The distinct roles that cognition and conation serve in our mental life can be understood as gaining support from the primary insight of *functionalism*, the theory in the philosophy of mind that identifies mental states neither in terms of the introspective features those states may possess, nor in terms of what they may be constituted by but in terms of their *functional*, or causal, role.[10] Insofar as we understand that distinction, we have a proposal for a theoretically respectable way of cashing out the metaphor of 'direction of fit.'[11] The 'cognitive/conative' distinction is, as we have seen, a distinction in causal role: cognitive states represent and manipulate information, conative states motivate, i.e. influence engagement. Yet, though functionalism provides theoretical cover to those philosophical psychologists who maintain this distinction, we ought not forget that this distinction in role is deeply rooted in our commonsense distinctions between 'reason' and 'feeling,' 'thought' and 'emotion,' and, indeed, 'belief' and 'desire.' The significance of this will bear considerably on what follows.

With this working account of the cognitivist and conativist theses of moral judgment in place, we can now go on to consider how the debate about moral judgment and the debate about moral motivation are related. For instance, we immediately see that conativism entails internalism: if moral judgments are themselves a function of an agent's motivational orientation, then it follows that a person who makes a sincere moral judgment necessarily

---

[10] Jerry Fodor puts the point as follows:

> The intuition that underlies functionalism is that what determines which kind a mental particular belongs to is its causal role in the mental life of the organism. Functional individuation is individuation in respect of aspects of causal role; for purposes of psychological theory construction, only its causes and effects are to count in determining which kind a mental particular belongs to. *Representations* (Cambridge: The MIT Press, 1981), 11.

[11] See Smith, "The Humean Theory of Motivation," 52-4, and *The Moral Problem*, 113-6.

has a certain motivational orientation towards the object of that judgment. If the moral judgment that killing innocents is wrong is a function of the disgust, horror, and aversion one has towards the killing of innocents then it follows that a person who sincerely judges that killing innocents is wrong is necessarily disposed to avoid, or eschew, the killing of innocents.

It is worth pausing for a moment to consider an implicit assumption that has been at work in the discussion so far, namely that cognitivism and conativism are mutually exclusive and jointly exhaustive options for interpreting the nature of moral judgment. Why assume this? Why not think that there is some possible hybrid of the two, or think that there is some altogether distinct—neither cognitive nor conative in nature—account of moral judgment? Some comments are in order. Firstly, the claim of mutual exclusivity is not meant to imply that the process of moral judgment is to be understood as exclusively a matter of cognition or exclusively a matter of conation. The most plausible and carefully formulated versions of both cognitivism and conativism make reference to both types of capacities in their accounts: moral judgment involves *both* cognition *and* conation. However, the question of the truth-value, or lack thereof, of moral judgments raises the issue of primacy: are conative capacities in the service of cognition or vice versa? Put another way, the question is whether our motivational states make correct moral descriptions possible,[12] or is it that defensible conative responses underlying moral judgments would be those made in the light of an accurate depiction of a *non-moral* reality? However we conceive it, we must say whether our moral judgments are truth-evaluable. If they are, cognition is the fundamental aspect of moral judgment, if not then conation is.

---

[12] As, for instance, 'sensibility theorists' maintain. See Stephen Darwall, Allan Gibbard, and Peter Railton, "Toward a *Fin de siècle* Ethics: Some Trends" *The Philosophical Review* 101 (1992): 115-189.

Secondly, with respect to cognitivism and conativism being jointly exhaustive, we can only rely on accepted taxonomies of the mental. It is assumed that cognitivism and conativism exhaust the metaethical possibilities since cognition and conation exhaust the psychic division of labor.[13] Unless there are some psychological capacities that are not properly understood as a species of either cognition or conation, the assumption of joint exhaustiveness would appear appropriate. I will assume this is so for the sake of the present discussion. But now notice what this permits us to say. The truth of conativism entails the falsity of cognitivism, and vice versa. Given the essential internalist character of conativism, we can conclude that if cognitivism is false then internalism is true: there is a necessary connection between moral judgment and motivation if moral judgments are not properly understood as depicting a moral reality.

Now what if internalism is false? Obviously we could conclude immediately that externalism is true. The immediacy of this inference is equally due to claims of mutual exclusivity and joint exhaustiveness. Yet, whereas these claims needed some elaboration in the case of moral judgment, their applicability in the motivational case is readily apparent: either there is a necessary connection between moral judgment and motivation or there is not. If internalism is true then externalism is false, and if externalism is true then internalism is false.[14]

---

[13] At least on the assumption being made here that the affective is a subspecies of the conative.

[14] There are, actually, some philosophers who deny the claim of joint exhaustiveness, for instance Jonathan Dancy in his *Moral Reasons* (Oxford: Blackwell, 1993), 6, 32-4; and Evan Simpson, "Between Internalism and Externalism in Ethics," *The Philosophical Quarterly* 49 (1999): 201-14. I do not believe that either attempt succeeds in showing that internalism and externalism are not jointly exhaustive. Simpson, in particular, provides a psychological-cum-semantic characterization of moral beliefs that appears to be true only if conativism were true, thereby entailing a conceptual connection between judgment and motivation. I cannot argue for this here, but a more comprehensive treatment of these issues is required. Dancy's account of motivation, of which he says that "there is a lot more about it that is internalist than there is externalist" (25) will be discussed at length later in this chapter.

But there is more. Given that internalism is entailed by conativism, by *modus tollens* we can conclude that if internalism is false then conativism is false. If the relationship between moral judgment and motivation is only contingent, then it cannot be true that moral judgments are a function of the moral agent's motivational orientation. But if conativism is false then cognitivism is true. That is, the contingency of the relationship between moral judgment and motivation implies that such judgments are beliefs about a moral reality. So the falsity of internalism entails the truth of both cognitivism and externalism, and the truth of externalism entails the truth of cognitivism and the falsity of internalism. Hence conativists are necessarily internalists and externalists are necessarily cognitivists. This, however, is as far as we can go by consideration of these theses in the abstract. For the claim that cognitivists must be externalists (and that internalists must be conativists), much more needs to be said.

Internalism, Cognitive Style

I will argue that cognitivist internalism (CI often hereafter) is false because it is incompatible with our commonsense understanding of wrongdoing and the explanations that we offer to account for it. These explanations also serve to provide categories of wrongdoing and wrongdoers—types of behavior and character that we take to be frequently instantiated. Since wrongdoing and wrongdoers are indisputable elements of the human experience, a theory that could not allow—not to mention explain—that experience should not face easy acceptance. Since wrongdoing and wrongdoers are, arguably, the main reasons that moral theory has practical significance, a theoretical approach that renders wrongdoing and wrongdoers impossible is an approach that is at best of academic interest only and, at worst, thoroughly perverse.

Yet, cognitivist internalism has been, and continues to be, thought by some philosophers to be not only viable but desirable. In order to show the confusion of this thinking we need to consider *how* CI *could* be true. If CI were true it would be so because one of two psychological theses were true. The first thesis is the claim that a certain class of cognitive states, namely evaluative beliefs, is intrinsically motivational. Mary's belief that killing children other than her own is wrong is an example of such a state. It is a belief about the killing of children other than her own and is, in the terminology we are using, a representational state. Furthermore, it represents the killing of children other than her own as being *wrong*, hence it is evaluative. The state aims at truth: Mary's belief is true if, and only if, the killing of children other than her own *is* wrong. As the thesis would have it, if Mary sincerely believes that killing children other than her own is wrong then she is motivated to avoid killing other mothers' children. This is so because, according to this theory, believing that killing children other than your own is wrong is to *both* represent such action as wrong *and* to have an aversion to it. Evaluative beliefs have, as it were, two directions of fit: mind-to-world and world-to-mind.

This contrasts with *non*-evaluative beliefs, such as the belief that Mary has killed her own child. This state is representational but involves no evaluation. Hence, this state is not motivational; its direction of fit is solely mind-to-world. Evaluative beliefs also contrast with conative states, such as the aversion Mary has towards killing children other than her own. This state is motivational but not representational;[15] its direction of fit is solely world-to-

---

[15] One must be careful here. This state, and perhaps most conative states, are *intentional*. That is, they have propositional content: Mary is in a state that is *about* the killing of children other than her own. However, Mary's *aversion* to such an act is not generally thought to be representational, nor is it truth-apt. Mary's aversion is an *attitude* that she has towards the proposition in question; it is her orientation to that proposition. It is, no doubt, of considerable significance that we can, and perhaps ought, to understand beliefs in a similar fashion. They, too, are attitudes towards propositions. Understood in this way, labeling beliefs as 'representational' is, at best, redundant, at worst, false. It might be suggested, however, that beliefs are representational in the sense that they represent the proposition in question as being true. The adequacy of such a response is, of course,

mind. Hence this taxonomy of psychological states includes three distinct kinds: beliefs, desires, and 'besires.'[16] Since moral judgments would, on this thesis, belong to the class of evaluative beliefs, moral judgments would be cognitive in nature. Since evaluative beliefs are construed as intrinsically motivational, internalism would apply.

The second, and different, thesis that supports CI is the claim that evaluative beliefs and motivational states, though functionally distinct, are nonetheless necessarily connected. On this view, the standard dichotomy between cognitive states, such as beliefs, and conative states, such as desires, holds fast: beliefs are representational but not motivational—they have only a mind-to-world direction of fit—and desires are motivational but not representational—they have only a world-to-mind direction of fit. This dichotomy holds even for evaluative beliefs. Hence, Mary's belief that killing children other than her own is wrong is intrinsically representational only, not intrinsically motivational. Nevertheless, this thesis maintains that Mary's belief (and all evaluative beliefs) are *relationally* motivational, and necessarily so. Put another way, Mary's belief is not *itself* motivational but *entails* motivation: Mary's (and our) motivational orientation is (in part, at least) beholden to the evaluative beliefs that she (and we) in fact has.

---

dependent upon the plausibility of truth's being an extra feature of a situation that we represent, rather than its being a 'result' of our *commitment* to a proposition's practical and/or theoretical significance. In this respect, the question of the appropriate understanding of the predicate '...is true' is quite similar to that of the predicate '...is wrong' and of moral predicates in general. I will not pursue these matters here and will conform to the standard practice of taking belief to be a cognitive (representational) state. It is worth noting, however, that the rudiments of a 'conative' account of belief can be found, I believe, in Hume's account of belief as a 'lively' idea.

[16] The term 'besire' was introduced by J.E.J. Altham in his "The Legacy of Emotivism," in *Fact, Science, and Morality: Essays on A.J. Ayer's Language, Truth, and Logic*, edited by Graham Macdonald and Crispin Wright (Oxford: Basil Blackwell, 1986), 275-88.

As for the use of 'desire', it is here meant as a stand in for conative states in general. This is a mere convenience and should not suggest any substantive thesis about those states, such as the view that all conative states are reducible to desires.

The difference between the two theses is significant. In its most well known version, the relational construal of CI is held to be defeasible.[17] In certain specified circumstances, the conceptual connection between evaluative judgment and motivation *can*, and may be *expected* to, break down. No such qualification is intended or even seems possible on the intrinsic construal of CI. For this reason it is important to keep these theses distinct. For simplicity I will refer to intrinsic CI and relational CI as 'strong internalism' and 'weak internalism,' respectively. Both, I will argue, are inadequate to the demands of our pre-theoretic moral practice.

It was said in the previous section that the distinction between cognitive and conative capacities was a 'commonplace' of philosophical psychology. It is so much of a commonplace in fact that it has been described as a "dogma."[18] The dogma, which has come to be known as 'the Humean theory of motivation,' in honor of its most prominent proponent,[19] has it that only conative capacities can be productive of action, and that cognitive capacities serve only an instrumental role—that of relating means to ends. What is, for our purposes, significant about this dogma is that the psychological theses that underlie it stand opposed to the theses that underlie both versions of CI. Underlying the Humean theory is the thesis that cognitive and conative states and capacities are distinct: cognitive states are representational while being motivationally inert, whereas conative states are non-representational yet motivationally efficacious. [20] Now obviously the basis of strong

---

[17] I have in mind Michael Smith's version of internalism. See, for instance, *The Moral Problem*, *op. cit.*, 61. His view will be discussed at length in the next chapter.

[18] Michael Smith, "The Humean Theory of Motivation," *op. cit.*, 31.

[19] David Hume, *A Treatise of Human Nature*, *op. cit.*, II. iii. 3.

[20] It has recently been suggested that the Humean about motivation ought not conceive of conative states as *necessarily* motivating, for to do so is to render Humeanism susceptible to attack from defenders of 'cognitive' theories of motivation. Since my aim here is not to defend Humeanism per se, but to critique cognitivism, I

internalism, the thesis that 'besires' have both cognitive and conative properties, runs afoul of this thesis. If there are such states as besires, then it is simply not true, as it would according to the Humean theory, that we need to make reference to a desire (or other conative state) in order to explain an agent's behavior. The presence of a besire, intrinsically motivational as it happens, is all we would need.

Clearly, then, either the 'dogma' of Humeanism is true or strong internalism is, but not both. I said that the theses underlying *both* versions of CI are opposed to the theses supporting the Humean theory. But how can this be so? Weak internalism *agrees* with Humeanism about the distinct intrinsic capacities of cognitive and conative states and thereby denies the existence of besires. Weak internalists, no less so than Humeans, take explanations of behavior to make essential reference to conative states, the only kind of state that has motivational force. Whither the problem?

The problem is that for the weak internalist, the essential reference to conation in the explanation of action is *guaranteed* by the reference to an evaluative belief. But this runs counter to the moral that Hume drew from the indispensability of conation to the production of action. From the distinctness of cognitive and conative states, Hume concluded that no connection between them could be discerned *a priori*.[21] There was nothing "contrary to reason" in a person being more highly motivated to pursue what she believed to be her harm than her good. Indeed, she need not be motivated by what she took to be her good at all. The belief that something will be to my harm, or that I can expect

---

will present Humeanism along the lines of the 'standard model.' The arguments against cognitivism presented here do not turn on this. For discussion of standard and non-standard models of Humean accounts of desire, see Steven Arkonovich "Defending Desire: Scanlon's Anti-Humeanism," *Philosophy and Phenomenological Research*, Vol. LXIII, No. 3 (2001): 499-519.

[21] David Brink suggests that claiming that the connection between judgment and motivation can only be known *a posteriori* is a way of being an externalist about motivation. See *Moral Realism and the Foundations of Ethics*, *op. cit.* 42.

some future good, is an evaluative belief. But, being beliefs, they have, according to both the weak internalist and the Humean, no motivational force. But since beliefs have no motivational force, there is nothing inconsistent or incoherent in a person having any or no motivational orientation with respect to them. Any claim about 'normal' associations between cognitive and conative states—such as desiring what is believed to be good and avoiding what is believed to be harmful—could be, at best, an empirical generalization. Weak (and strong) internalism, however, claims the connection between these states to be necessary. It is therefore incompatible with the spirit (and the logic), if not the letter, of Humeanism.[22]

It is important to see that, quite in contrast to the situation of the cognitivist internalist described above, both the conativist and the externalist can (and must) take Humeanism to be correct. Hume, of course, was himself a conativist, so how could both externalism and conativism be consistent with the Humean theory of motivation? What drives the wedge between the conativist and the externalist is the issue of the putative practicality of ethics. The Humean theory of motivation is itself silent over the conativist/cognitivist dispute. It is only when coupled with the thesis that morality is essentially practical, i.e. has an influence on actions, that is to say is motivational, can Humeanism about motivation support the conclusion that moral value is an expression of conative states and not discerned by cognitive states. Hence externalists, who deny that morality is essentially practical, can consistently maintain cognitivism about moral judgment and Humeanism about motivation. Cognitivist internalists, on the other hand, attempt to marry a representational construal of moral judgments with a practical construal of morality and must therefore view the

---

[22] Though see Jonathan Dancy, "Why There is Really No Such Thing as the Theory of Motivation," *Proceedings of the Aristotelian Society* (1995), 1-18.

relationship between judgment and motivation far more intimately than would the Humean about motivation. Indeed, the connection is so intimate that the Humean theory is rendered false (in the case of the strong internalist) or irrelevant (in the case of the weak internalist). But being dismissive about Humeanism on motivation is akin to theoretical suicide. The problem is that our pre-theoretic understanding of our moral life—and, most importantly, the concepts and categories that comprise it—is predicated on the truth of the Humean view. It is a dogma in philosophical psychology precisely because it is a pillar of our commonsense, belief-desire psychology: beliefs represent, desires motivate, and no belief guarantees any desire. If this picture is false then our commonsense understanding of much of human behavior—moral and otherwise—is radically misconceived.

The remainder of this chapter will consider in some detail the work on wrongdoing of the leading contemporary strong internalists, notably John McDowell, David McNaughton, and Jonathan Dancy. We should expect to find, however, nothing too different from what we saw in the previous chapter's discussion of Socrates' account of wrongdoing in *Protagoras*. That is because, as was argued there, Socrates' internalism, captured in his claim that "no one goes willingly toward the bad or what he believes to be bad," is based upon the thesis that motivational states are properly construed as evaluative beliefs. We saw, too, that such a construal forces Socrates to offer the ignorance thesis of wrongdoing, a thesis that we argued is properly understood as a 'negligence' thesis. But the negligence thesis is actually not available to Socrates. If negligence is to be distinguished from ignorance it can only be so on the basis of a motivational difference. But is precisely this that is beyond the scope of a strong internalist: if moral judgments are intrinsically motivational then the lack of motivation to do what is right or best can only be due to the absence of the appropriate

evaluative belief. In other words, it is moral ignorance that is the cause of wrongdoing. Since true ignorance is not blameworthy, the result is the denial of wrongdoing altogether.

We now have a diagnosis of where Socrates went wrong. By attributing motivational efficacy to cognitive states like evaluative beliefs, Socrates would have denied the Humean theory of motivation. Without being a Humean about motivation an attempt to explain wrongdoing will prove hopeless. The common moral categories of negligence, weakness, wickedness, indifference, and evil all presuppose a Humean understanding of psychology and motivation, for they explicitly permit and/or require the divergence of cognition and conation. This point was apparently not lost on Socrates. Indeed, as we saw earlier, his discussion of wrongdoing begins with him drawing a line between himself and the common run of men. "Most people," Socrates claims, think that "while knowledge is often present in a man, what rules him is not knowledge but rather anything else—sometimes desire, sometimes pleasure, sometimes pain, at other times love, often fear; they think of his knowledge as being utterly dragged around by all these other things as if it were a slave."[23] Commonsense, as Socrates sees it, takes desire and emotion and the like to be distinct from the grasping of truth and that possessing the truth, or simply believing oneself to be in possession of it, and is no guarantee that a particular motivational orientation will be present or produced. Commonsense, it would seem, is rather Humean in its understanding of psychological capacities and their role in the explanation of behavior. Outside a Humean understanding of psychology, weakness, wickedness, wanton cruelty, and the rest are simply inexplicable. The moral seems quite clear: deny a Humean (i.e. folk) psychology and you deny the very phenomena that a moral theory is supposed to explain.

---

[23] Compare Hume: "Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them." *Treatise of Human Nature*, II. iii. 3. (415).

The New Socratism

> The internalist realist rejects the terms within which his opponents are conducting the debate. He denies that cognitive states are incapable of moving an agent to act. He can therefore allow that the belief that he is morally required to act is sufficient to move the agent to act, with out assistance from a quite different kind of state, a desire. Since he is not saddled with a picture of the mind in which there is drawn a sharp distinction between passive, cognitive states and active, motivating desires, the internalist realist sensibly bypasses the dispute between non-cognitivists and externalist realists as to the side of that fence on which moral commitments are to fall.[24]

So says David McNaughton in his spirited defense of the cognitivist internalist position in his aptly titled work *Moral Vision.* Here McNaughton is firmly identifying himself with what we have called the strong internalist position. His internalism is strong in that he countenances the positing of cognitive states that are intrinsically motivating—besires. It is such states that the strong cognitive internalist—in McNaughton's terminology, an 'internalist realist'—takes as underlying moral judgment. McNaughton spells it out in the following passage.

> To be aware of a moral requirement is, according to the realist, to have a conception of the situation as demanding a response. Yet to conceive of a situation as demanding a response, as requiring one to do something, is to be in a state whose direction of fit is: the world must fit this state. The requirement will only be satisfied if the agent changes the world to fit it. But the realist also wishes to insist that the agent's conception of the situation is purely cognitive. That is, the agent has a *belief* that he is morally required to act and so his state must have the direction of fit: this state must fit the world. For his belief will only be correct if it fits the world, if it accurately reflects the way things are. If he becomes convinced that things are not, morally speaking, as he believes them to be, then he must give up his moral belief. He is committed, therefore, to the claim that the awareness of a moral requirement is a state which must be thought of, Janus-like, as having directions of fit facing both ways. The agent's conception reveals to him both that the world is a certain way and that he must change it.[25]

---

[24] David McNaughton, *Moral Vision* (Oxford & Cambridge: Blackwell Publishers, 1988), 49.

[25] *Moral Vision, op. cit.*, 109.

McNaughton's views about moral judgment are, as he acknowledges, inspired by the work of John McDowell.[26] In a number of articles, McDowell has developed a realist position with the intent to resuscitate the idea, found both in Plato (and Socrates) and Aristotle, that 'virtue is knowledge.'[27] The virtuous person, according to McDowell, is one who "conceives" of her situation in a particular way, a way that provides her with all the reason (motivation, in McDowell's sense) she needs to act. Saying this, however, renders the Humean claim that an explanation of action requires reference to some distinct, or 'extra,' motivational state superfluous. Taking the conception of one's situation, or the recognition of one's reasons for acting, to be sufficient in and of itself to move the agent has become an increasingly popular position in discussions of moral psychology and theories of practical reason. This cognitivist account of motivation gained momentum with Thomas Nagel's distinction between "motivated" and "unmotivated" desires, the former being consequent on considerations or the recognition of reasons that are themselves motivational. Hence, as Nagel, and later McDowell (and a host of others) see it, the ascription of a desire (or some conative state) to an agent that has acted is always necessary because of the *logic* of psychological explanations of action. Given that an agent was motivated to act by her conception of her situation, it is appropriate to say that she 'desired' to act in that way. Yet this is not a substantive concession. There is no suggestion here that we need think of the

---

[26] Particularly relevant here are McDowell's "Are Moral Requirements Hypothetical Imperatives?", *Proceedings of the Aristotelian Society*, Supplementary Volume, 1978, 13-29, and "Virtue and Reason," *The Monist*, Vol. 62 (1979): 331-50. Reprinted in Roger Crisp and Michael Slote, eds., *Virtue Ethics* (Oxford: Oxford University Press: 1997): 141-62. Subsequent page references to the latter.

[27] "Virtue and Reason," ibid., 331.

ascribed desire as a non-cognitive, motivational state which was required to move the agent; her conception of the situation was sufficient for that.[28] As T. M. Scanlon puts the point,

> a rational person who judges there to be compelling reason to do A normally forms the intention to do A, and this judgment is sufficient explanation of that intention and of the agent's acting on it...There is no need to invoke an additional form of motivation beyond the judgment and the reasons it recognizes, some further force to, as it were, get the limbs in motion.[29]

However, the question we are asking is whether such a view of cognitive states can provide intuitively satisfying explanations of wrongdoing. The claim here is that it can't. If this is right then any theorist that employs such a psychology has a deeply problematic account on offer. They will be unable to divorce the recognition of a moral reason or the conception of a situation in a morally significant way from the motivation to act accordingly. As such, understanding wrongdoing will be beyond their scope. For convenience, though, I will focus the discussion in the remainder of this chapter on the views of McNaughton and McDowell and some related criticisms of them.

The virtuous person, according to McDowell, is one who is adept at perceiving the morally "salient" features of any given circumstance where "the relevant notion of salience cannot be understood except in terms of seeing something as a reason for acting which *silences* all others."[30] Hence, as McDowell (and McNaughton) would have it, the agent who perceives what ought to be done in a given situation has no reason to act any other way. But

---

[28] Thomas Nagel, *The Possibility of Altruism* (Princeton: Princeton University Press, 1970), 29-30. McDowell endorses this line of argument in "Are Moral Requirements Hypothetical Imperatives?", 15-6. For an extended discussion of the significance of this distinction for 'belief-desire', or 'Humean' accounts of explanation and justification, see G.F. Schueler, *Desire: Its Role in Practical Reason and the Explanation of Action* (Cambridge: MIT Press, 1995). For a discussion of cognitive theories of motivation that questions whether the Nagel/McDowell conception of desires is to be non-cognitively construed, see Dancy, *Moral Reasons*, ch. 1-2.

[29] T. M. Scanlon, *What We Owe to Each Other* (Cambridge: The Belknap Press of Harvard University Press, 1998), 33-4.

[30] "Virtue and Reason," 158. Emphasis added.

this, of course, implies that those who *do* have reason—those who are in some sense motivated to act as they ought not—fail to perceive the morally salient features of their context. As McDowell says, "one cannot share a virtuous person's view of the situation in which it seems to him that virtue requires some action, but see no reason to act in that way."[31] This follows inevitably from the view that moral judgments are cognitive states that are essentially motivational, that is, besires. If someone does wrong, on this psychology, she inevitably lacked the appropriate besire. It was argued in the previous chapter that it was precisely this psychological commitment that supported Socratic internalism, leading ineluctably to his conceptually flawed ignorance thesis. There would seem no reason to think McDowell and McNaughton can evade a similar fate. Nevertheless, both discuss wrongdoing in the presentation of their views, attempting to assuage concerns about this putative theoretical weakness, with McNaughton devoting considerable attention to it. I want to turn to that discussion now. Before I do, however, there is a point that warrants attention, namely the proper extension of besires. We have, for mere convenience, equated besires with 'evaluative beliefs,' yet evaluative beliefs would seem to be a wider class of judgments than merely *moral* beliefs. There are, for instance, judgments of prudence, aesthetic evaluation, and mere preference or taste, such as my belief that meatloaf and potato pie is a wonderful meal on a winter night. Are *all* such beliefs intrinsically motivational or only the moral ones? Alternatively, we can ask whether *all* motivation is properly understood as evaluative beliefs or only moral motivation?

The issue here is whether the strong cognitivist internalist is offering what McNaughton calls "mixed" or "global" versions of a "cognitivist theory of motivation."[32] McNaughton is

---

[31] "Are Moral Requirements Hypothetical Imperatives?", ibid., 26.

[32] *Moral Vision*, ibid., 106-7, 127-ff.

himself at times rather non-committal on this issue, but he does concede that "there is always something unsatisfying about a mixed theory; considerations of simplicity lead one to prefer a system where all motivation is of the same kind."[33] This seems right, but we can add to this by pointing out that there seems no non-arbitrary way of distinguishing between motivational states that are susceptible to cognitive analyses and those that are not. Consider, for instance, the vacuity of grounding the distinction by reference to the 'concept' *good* exemplified in the following passage from Colin McGinn.

> Besides, [conativism] does not avoid the problem of saying what sort of property goodness is, since the attitudes it invokes must always contain the concept *good* as part of their content or else fail to constitute a genuinely moral evaluation. The emotion I express by saying 'That is good' will be an emotion of approval, but that is an emotion in whose content the concept of goodness occurs—it is a feeling *that* something is good. If we now try explain this concept by appeal to other emotions and attitudes expressed by moral words, we generate an infinite regress, since these further psychological states will have to have the concept of goodness built into them too, on pain of not constituting *moral* attitudes.[34]

There is much that one could complain about in such an argument. From the perspective of the commonsense/Humean psychology, which distinguishes between cognition and conation, the claim that the emotion expressed by saying 'That is good' will be "an emotion in whose content the concept that goodness occurs—it is a feeling *that* something is good" is a barbarism. One has feelings *about* something thought to be good (or, simply, that *is* good): one can be angry or pleased, frustrated or sad, sick or ecstatic that this is good. But this is not to say that there is a feeling *that this is good.* There is no such feeling. Nevertheless, the real difficulty for McGinn is in determining which attitudes have the concept of goodness "built into them" and those that do not.

---

[33] *Moral Vision*, 130.

[34] Colin McGinn, *Ethics, Evil, and Fiction* (Oxford: Clarendon Press, 1997), 8. Emphasis in the original. I have replaced McGinn's use of 'non-cognitivism' with 'conativism.'

Consider Jim's motivation to keep his promise to his wife to quit smoking. Jim believes that keeping his promises is a good thing to do. On the thesis in question, then, Jim's motivation to keep his promise is a moral motivation. But what of his motivation to continue smoking? To an adherent of a mixed cognitive theory of motivation like McGinn, this motivation is not moral since it does not have the concept of goodness as a component of its content. But why not? Why isn't Jim's motivation to smoke properly expressed by the statement that smoking is good? Imagine that Jim's desire for smoking is strong and that it gives him considerable satisfaction. He is quite aware of the increased probability of acquiring some smoking related disease and the unfortunate smell—the considerations that lead his wife to want him to quit. But he is also aware that statistics, being the funny things they are, are not always so clear-cut. For instance, Jim has been impressed by reading that over 80 percent of those who smoke *do not* acquire smoking related diseases, so he thinks that the probability of smoking causing him harm is not that great. Furthermore, he believes that the irritability that quitting will cause and the long-term resentment that will ensue will be unhealthy for his relationship. He sees quitting—something he has an aversion to doing—as a bad thing. He reasons that the risks to his relationship if he quits are just as significant as those to his health if he continues smoking. Since he cherishes the relationship he has with his wife he concludes that, for him—and, by extension, for them—smoking is good.

I see no principled, non-arbitrary, and non-question-begging way for the cognitive theorist to distinguish moral motivation from non-moral motivation: if we are to countenance motivational states as being analyzed as besires, we should say that all are such. It would be good if we could leave this issue here; however, mixed cognitive accounts will

resurface later in the chapter. For now, however, let us assess the merits of the strong internalist's explanations of wrongdoing while taking the scope of her theory to be global.

The New Socratism and Wrongdoing

The argument being offered here is that cognitivist internalism is incompatible with our commonsense conceptions of wrongdoing and their explanation. This type of argument has had relevance for moral philosophy since it began; in the case of explaining wrongdoing, fidelity to commonsense seems to be at a premium. As McNaughton says "A theory that cannot give a satisfying account of how our moral convictions find their expression in action must be rejected. But our conviction that morality is action-guiding must be balanced against the common observation that people can fail to be motivated by moral considerations. An acceptable theory must explain not only how awareness of a moral demand can move an agent to act, but also how it can fail to do so."[35] Indeed, it is among the primary arguments offered by both McNaughton and McDowell to supplant the Humean theory of motivation with the cognitive theory that the former cannot give a satisfying account of the "phenomenology" of the moral life, while the latter can.[36] Yet this is, I maintain, precisely what the strong internalist cannot do. The strong internalist's account of wrongdoing sounds no better to the moral ear than the Socratic account of *Protagoras*.

Consider an explanation of someone's wrongdoing by way of the attribution of moral weakness. This, as we have said, is perhaps the most frequently used explanation of wrongdoing from the perspective of common moral practice. Indeed, weakness is as

---

[35] *Moral Vision*, 118.

[36] *Moral Vision*, 48-50, 114-5; "Virtue and Reason," 160-1.

frequently self-ascribed as offered from the third-person perspective. It is also, undoubtedly, the most discussed explanation of wrongdoing to be found in the philosophical literature. Discussions of weakness, though, like discussions of anything else, are only as adequate as they are faithful to the actual phenomena. If a philosopher misdescribes her subject her subsequent discussion of it will be inevitably unsatisfying. This problem is readily apparent in "the solution" McNaughton offers to weakness: "The solution, on the global theory, to the problem of moral weakness, and weakness of will in general, lies in the possibility of competing conceptions."[37] According to the global conception of cognitive motivation, all motivation is analyzable as evaluative belief. Hence the conflicted nature of the weak agent is understood in terms of a battle of evaluative judgments: Jim sees one course of action, say keeping his promise, as the loyal and honest thing to do, and sees another, say continuing to smoke, as being essential to an honest and healthy emotional engagement with himself and his spouse. Crudely put, Jim believes of two incompatible courses of action that they are both right.

We should not find this account of the phenomenon of weakness to our liking. Weakness, as it is commonly understood, is not the problem of deciding which of two morally acceptable (or required) but incompatible courses of action to take. There is, no doubt, such a difficulty which moral agents may frequently face, but it is not a case of moral weakness (or even weakness of will). How could it be if whichever course of action the agent pursues is one that he thinks is morally right? Weakness is an explanation of *wrongdoing*. In the construal of the phenomenon by the lights of the strong internalist, wrongdoing drops out.

---

[37] *Moral Vision*, 130.

One might object that we are being too quick with the strong internalist here. The competition between the different conceptions is supposed to be the starting point, not the end point for this theorist. "On the cognitivist approach the aim of practical reasoning is to organize these competing conceptions into an overall picture in which the various considerations find their proper place."[38] Our criticism fails to take account of the role of deliberation. Jim's initial judgments that both keeping his promise and continuing to smoke are morally defensible are, we might say, *prima facie* judgments. What is wanted is an evaluation of the situation 'all things considered' that would, presumably, *rank* the initial *prima facie* judgments.[39] Then Jim would have an overall conception of the situation that would, presumably, render the call to keep his promise as *the* right course to take. Now relative to that conception, the continuation of smoking could only be seen as an unacceptable alternative. If, then, Jim were to choose to smoke he would be doing something that, by his own lights, he took to be wrong. Given that Jim continues to smoke away, how is this to be explained?

McNaughton's answer to this seems to give back the ground just gained. We wanted the morally weak agent to do wrong on pain of not acting weakly. The strong internalist (globally construed) attempts to provide for this by appeal to an "overall" conception of the situation which provides for moral ranking. Given such a conception, we can imagine the agent being motivated to do the morally required act (relative the conception) since conceiving of an act as morally required is intrinsically motivating. But how are we to explain the eventual performance of *any* action other than a right one? To be motivated to perform any action, according to the cognitive theory, is to conceive of it in a positive light.

---

[38] *Moral Vision*, 130.

[39] See Donald Davidson's "How is Weakness of the Will Possible?" *op. cit.*

But it should be just such evaluative considerations that are "silenced" in the overall conception. The only answer that seems open to the strong internalist is the very answer McNaughton gives:

> Although the competing and more limited conceptions are taken up into the overall one, they are still available to the agent. Even after he has formed a judgment as to what is best, it may still be possible to view the situation from just one point of view, ignoring the aspects that would have to be fitted into a rounded picture...The harder he finds it to maintain the overall conception, and the more prone he is to fall back into the more limited conception, the more attracted he will be to a course he judges to be the worse.[40]

The problem with this line of analysis is not that the agent is motivated to do (and eventually performs) an act that he judges to be worse, but rather, from the 'inside,' he is motivated to perform the action he judges to be best. If the wrongdoing agent has 'fallen back' into a more limited conception then he is, by his own reckoning, pursuing a course of action that he thinks is appropriate. But we have again lost wrongdoing in the process, at least in the phenomenological sense that is an element of our commonsense picture of weakness. For that picture has agents performing the action they believe to be the worse *as they do it.* It is not simply by regret, repentance, and guilt that we demarcate the weak from other forms of wrongdoing, such as wickedness and indifference. We also want to distinguish weakness from negligence, and for that we need more than post-action assessment. We ought to allow for the possibility that the conflict that marks weakness is present throughout. The great indignity of weakness is the inability to hide it from ourselves: we are aware of our fall as it happens.

There is a further problem that the "competing conceptions" approach to weakness poses, a problem that threatens the very cogency of the global strong internalism that underwrites it. In discussing how the agent may fall back into a more limited conception, a

---

[40] *Moral Vision*, 130-1.

conception whereby the agent will be motivated to perform an action that is judged worse by the lights of the overall conception, McNaughton asks a question that is particularly to the point: "We may wonder why it should be hard to maintain the overall conception; if I have reached a view in which all the things to be said for and against a course of action have found their proper place, how can I then fall back into a more partial view?"[41] Indeed, how could one do so? McNaughton's response appears, on the surface anyway, to be mere hand waiving: "We need to remind ourselves that the ability to see the world as the virtuous person sees it is not one that is easily attained...in trying circumstances it is all too easy to fall back on less demanding conceptions."[42] True or not, this goes nowhere towards explaining *why* it is so difficult to maintain the virtuous conception and so easy to fall out of it. Yet an explanation is to be found, I think, in an example McNaughton gives to make his point clearer. It is, however, an explanation that is not in the global strong internalist's interests to give.

McNaughton asks us to consider a "naturally somewhat cowardly person" who finds himself in a situation where he can aid a friend by some act of bravery. From the perspective of the person's personal safety and well-being, the action is unattractive and of little positive value. Yet from the perspective of being a loyal and reliable friend, the action has much to recommend it. Indeed, if the person finds his cowardice shameful he may well view the performance of a courageous act to be an opportunity for long-desired change, the chance to become a better man, one he could be proud of. As McNaughton states, "In reaching that conclusion [he] may be guided by a conception of the kind of person [he] would like to be: a person loyal to his friends, adventurous and free of fear." But such

---

[41] *Moral Vision*, 131.

[42] *Moral Vision*, 131.

conceptions are not easy to maintain: our habitual ways of reacting to situations die hard, and our characters are not easily fought against. Being the person he is, our agent is one to put the avoidance of danger at a premium. It is likely then that when the time comes to act, he will not see an opportunity to grow in stature but a threat to his very being. In such a circumstance, we may say that the agent acted as he did from weakness.

This analysis of the moral agent and his difficulties is a problematic one for the strong internalist. The cognitivist wants to maintain, contra the conativist, that moral judgments are a species of cognitive judgment, a matter of depicting the moral realm as it is. Furthermore, being internalist about motivation, this cognitivist claims that judgments of value are intrinsically motivating, requiring no help from motivational states that are understood non-cognitively. Indeed, on the global construal of strong internalism there are no non-cognitive motivational states. But notice what we are told in McNaughton's example. The conception of the brave act as the one that ought to be pursued is "guided" by a conception of the kind of person the agent "would like to be." The agent is here depicted as wanting to be of a certain type, to be different than he actually is. Likewise, the agent's "habits of life and childhood training" encourage the assessment of brave actions as unduly dangerous and to be avoided. But are such motivations available to the strong internalist? Motivation, on this view, is consequent on the moral conception of the situation, not a condition of that conception (as the conativist would have it). But this implies that the conception that holds the brave act to be best is itself motivated by the conception that being a certain type of person is best (just as the alternative conception that views self-protective behavior as best is motivated by the conception of a person who avoids danger is best). And here, it seems, the strong internalist is faced with a dilemma. For, according to the first horn of the dilemma, if the explanation of why an agent might fall back on more restricted conceptions of a situation

is due to the agent's view that the more restricted conception is the one that ought to be adopted, we have again lost wrongdoing from the perspective of the agent. As far as the agent is concerned, the more restricted conception is to be favored over the more inclusive: it is judged to be of greater value than the overall conception, otherwise what would explain its greater appeal and motivational efficacy? Since being a person who looks after his own skin first is the best type of person by the agent's lights, we can explain why the conception that would highly value self-protective behavior is maintained over the "overall" conception that ranks bravery as best. But why would such an agent think he was doing wrong as he ran away?

The second horn of the dilemma attacks the integrity of strong internalism where it lives. If what drives agents to this or that conception is *not* construed as an evaluative judgment in its own right, it must be construed non-cognitively. But this is to give up the game for the strong internalist (globally construed, at any rate). If the agent's conception that bravery and loyalty to friends is best is guided by the kind of person he wants to be and the conception that places self-protection at the pinnacle is driven by his fears and insecurities, built up over a lifetime, then the way moral agents 'conceive' the world would seem consequent on their motivations, and not the other way round. Appealing to the agent's aspirations, hopes, fears, and habits is actually grist for the conativist's mill. Just as we saw in the case of Socrates, if an agent's behavior is to be construed as blameworthy, appeal must eventually be made to that agent's motivations. But if motivations are conceptually bound up with evaluative judgments then no such appeal can be made. The attempt to avoid this predicament leaves the strong internalist in danger of allowing evaluative judgment to be a function of motivation. But this is to save cognitivist

internalism by rejecting it for conativism. The global construal of strong internalism is not sustainable.

Some Cognitivist Criticism: Milo

The difficulties that global strong internalism faces in attempting to explain wrongdoing have not gone unnoticed, even by strong internalists themselves. Insofar as we accept that wrongdoing occurs, we want the psychological resources to explain it. We rob ourselves, however, of just such resources if we understand all motivation to be a function of evaluative judgment where such judgments are construed cognitively. This has inevitably led to attempts to modify the cognitive theory of motivation, to temper it in ways that will allow for a more intuitive approach to the problem of wrongdoing. There are a few proposals to this effect to be found in the literature; we need to consider them to see whether even a modified strong internalism is viable approach to the psychology of morality.

In his paper "Virtue, Knowledge, and Wickedness,"[43] Ronald Milo objects to the global line of McDowell and McNaughton and urges that a distinction be drawn between the idea that moral convictions are *intrinsically* motivating and that they are *necessarily* motivating. In speaking of the McDowell/McNaughton line, he says

> They contend that the virtuous person's moral conviction motivates her virtuous behavior without the addition of a *separate* desire to do what is morally called for—which is to say that this moral conviction is *intrinsically* motivating. And I do not dispute this point. What I deny is that no one else can have this same moral conviction without being similarly motivated. That is to say, I deny that this moral conviction is *necessarily* motivating—motivating for everyone who has it.[44]

---

[43] Ronald Milo, "Virtue, Knowledge, and Wickedness," in *Virtue and Vice*, edited by Ellen Frankel Paul, Fred D. Miller, Jr., and Jeffrey Paul, (Cambridge: Cambridge University Press, 1998), 196-215. The collection is a republication, with an introduction and index, of the semiannual journal *Social Philosophy & Policy*, Volume 15, Number 1.

[44] Ibid., 200. Emphasis in the original.

Milo admits in a footnote that after a presentation of this paper audience members, including Robert Audi and David Copp, challenged the distinction between the intrinsically and necessarily motivating conviction (moral or otherwise, presumably). I must confess that I also find the distinction in the present context less than obvious. If a moral belief is, in and of itself, sufficient for motivation then, *ceteris paribus*, it should be sufficient for motivation regardless of who holds the belief. Of course, the crucial step here is to sufficiently spell out the *ceteris paribus* clause so that we can be sure if it has been violated or not. And it is here, I think, that Milo makes a mistake. For Milo, it is clear that we can say that "for a virtuous person the conception that an act is required by virtue is intrinsically motivating; but for the wicked it is not."[45] Milo wants to be able to say this because he wants to deny, rightly, the claim that a wicked person cannot share the virtuous person's moral conception, a claim that McDowell and McNaughton endorse.[46] For as they would have it, a wicked person is one with a 'distorted' sense of moral propriety, and it is this difference in moral judgment that explains the difference in his behavior.

Where Milo goes wrong becomes clear in his attempt to make explicit the difference between the virtuous and the wicked, the difference that he claims accounts for the differing motivational efficacy of a moral conception. Milo claims that

> For those that love virtue (the virtuous), [the thought that an act is required by virtue] is *intrinsically* motivating. This is just to say that for the virtuous, this thought is sufficient to motivate them to act virtuously, without needing as a precondition some additional, non-cognitively understood desire to do what virtue requires. Again, however, this seems compatible with holding that for other people—the wicked, who do not love virtue—this same thought fails to be motivating. Thus, although the thought that an act is required by virtue is *intrinsically* motivating for those who are motivated by it (the virtuous), it is not *necessarily* motivating (since the wicked fail to be motivated by this thought). It is

[45] "Virtue, Knowledge, and Wickedness," 201.

[46] *Moral Vision*, 140-4; "Are Moral Requirements Hypothetical Imperatives?", 23, 26.

> important to note that the internalists about the connection between moral beliefs and motivation whom I am discussing in this essay hold that moral beliefs are *both* intrinsically and necessarily motivating. This holds true of internalist moral realists like McDowell, as well as [conativists].[47]

Milo's mistake should be obvious immediately. What is apparently doing the work required to mute the motivational force of the wicked agent's moral beliefs is his lack of love of virtue. The virtuous, who 'love virtue'—apparently as a matter of definition—find their moral beliefs to be intrinsically motivating. But why—more importantly, how—could *love* have this influence? The (global) strong internalist claims that all motivational states are properly analyzed as evaluative beliefs. If a state is not a cognitive state of that kind then it simply is not motivational. For strong internalists, there is no need to posit an additional, non-cognitive, state to account for motivation because no such state is motivational. So, what's love got to do with it? It seems clear that there are only two possibilities. The first is that love is a non-cognitive state and, by the strong internalist's lights, is not motivational. Perhaps it is a purely affective state with some phenomenological quality. In that case, then, the virtuous person's love of virtue and the wicked person's lack of love (or even disdain) have no bearing on the issue of the motivational force of the moral beliefs themselves. If, as the strong internalists maintain, moral beliefs are intrinsically motivating (as Milo apparently concedes) then they are necessarily motivating as well. If Milo thinks of love as a non-cognitive state, his argument has no force.

The second possibility is that love is an evaluative belief, and is, therefore, motivational. In that case, the presence or absence of love will have motivational consequences. The lover of virtue (the virtuous) will be motivated to perform virtuous acts whereas those who have disdain for virtue (the wicked) will not (they will likely have an aversion to such acts). There

---

[47] "Virtue, Knowledge, and Wickedness," 201, n.12, emphasis in the original.

is, no doubt, much truth in this. We want to preserve the idea that wicked people *want* to do wrong—that is just what it *is* to be wicked. I also think, with Milo, that strong internalists have great difficulty allowing for this possibility (if they can at all), which is one reason their views should be rejected. But it is not clear that Milo's argument, as given, is the way to show this.

Assuming with the strong internalist, as Milo apparently does, that moral beliefs are themselves motivational, making additional reference to the love (and disdain) for virtue becomes particularly problematic. The strong internalist maintains that an agent who judges, say, that bombing innocent civilians is wrong, necessarily has an aversion to such activity. Imagine a person who is, as the strong internalist would have it, appropriately motivated by all their moral judgments. Is such a person virtuous? He may not be if, as Milo implies by his contextual definition, the virtuous person is a *lover* of virtue. Add to this the claim that love is an evaluative belief and we have a view of the virtuous person as one who judges being virtuous to be good. Put another way, the virtuous person is one who judges that he ought to do what he ought to do, and is not merely one who does what he ought. Perhaps this is part of our conception of the virtuous, enabling us to distinguish them from the merely good, for the former, as opposed to the latter, are *doubly* motivated to do as they ought. It is not clear to me whether strong internalists such as McNaughton and McDowell hold this account of the virtuous person, that is whether they think the virtuous person must be a lover of virtue, but I do not think it is particularly troubling if they do, no more so, anyway, than their initial view that moral judgments are to be construed as beliefs that are intrinsically motivating. But I do think that it is not a view that Milo should adhere to, at least given his argumentative aims.

Milo wants to defend the possibility of explaining wrongdoing by reference to the wickedness of the agent, an explanation that he, rightly, thinks is unavailable to the strong internalist (globally construed). But notice what Milo becomes committed to when he combines the claim that moral beliefs are intrinsically motivating (which he claims he does not want to deny) and the claim that the virtuous are lovers of virtue (when love is understood as an intrinsically motivating evaluative belief). A wicked person, according to Milo and commonsense, is a person who is capable of making the very same moral judgments that a good (or virtuous) person can, but is motivated to do what he judges wrong. But if the wicked person is marked by his disdain for virtue, then the wicked person is properly construed as someone who, on the one hand, is motivated to do the right thing and, on the other, averse to being motivated to do the right thing. Taking our previous example, the wicked person judges that bombing innocent civilians is wrong and, since this judgment is intrinsically motivating, he is averse to so acting. But, further, this person is one who disdains virtue, hence he has an aversion to being averse to doing the wrong thing. But this would render a picture of the wicked agent as conflicted—differently, no doubt than the conflict present in the weak, but conflicted nevertheless—and this does not seem right.[48] We want a class of agents who can knowingly do wrong *without* having to fight through an aversion to so acting. An important element to the commonsense understanding of wickedness is the *lack* of conflict. We want such an explanation of wrongdoing available to us, if we can have it, but it appears that the psychology we are considering will not permit it.

[48] This is not to say that there are no agents so described, or that this is not a plausible explanation of some wrongdoing. What is being objected to is the conclusion that this is all there is to the category of wickedness. Another plausible (and far simpler) explanation of wrongdoing is that the agent *wanted* to do the wrong thing and felt no inhibitions about doing it. This explanation is ruled out on the current proposal.

To be fair, Milo speaks of the wicked agent as one whose moral beliefs "fail" to be motivating. Milo, it is apparent, does not construe the wicked as conflicted in this way but as those whose normally motivating beliefs are robbed of their motivational force by an agent's evaluations of morality itself. But Milo provides no argument as to why we should see the evaluation of morality itself as robbing moral judgments of their intrinsic motivation as opposed to overpowering them. In the absence of such an argument, coupled with the assumption that moral judgements are intrinsically motivating, we are given no reason to think that such judgments are not *necessarily* motivating. And this, as we have seen, denies us a plausible explanation of some cases of wrongdoing.

It is worth pausing on this point for a moment. We could attempt to construct an argument showing that an agent's evaluative beliefs about morality itself not only supercede but render ineffectual his first-order moral judgments, hence preserving the commonsense explanation of wrongdoing that makes appeal to the intentional, and unconflicted, wrongdoer. But it is not at all obvious that we should or that we even need to do so. If moral beliefs are intrinsically motivating then, for the psychologically well-adjusted at any rate, we have all we need to explain virtuous behavior (providing, of course, that those moral beliefs are true). Love of virtue need have nothing to do with it. Likewise, we want to allow for the (occasional) intentional, non-conflicted wrongdoing by those who do *not* disdain virtue. Indeed, I think it is essential that we make room for the countless acts of 'minor' or 'local' wickedness that occur everyday. People frequently do and say that which they know would be wrong, intentionally, without necessarily having disdain for virtue. This is, sadly, perhaps most true in the case of our more intimate associations, those we have with our closest friends and family. We often say things to our spouses and family members that we know will cut them deeply, striking them 'right where it hurts.' We know what buttons to

push with these people, know exactly how to get them to acknowledge our anger, resentment, and frustration. With clear sight as to what is wrong we do it willfully, and this is wicked, in its way. But to explain our shamefulness must we posit disdain for virtue? Not at all. We only need acknowledge that our moral judgments are not necessarily motivating. But to do this is to deny internalism (however it is construed, strong or weak). And it is this, really, that Milo wishes to do. Milo really doesn't mean to concede to the internalist the claim that moral judgments are intrinsically motivating. He is, after all, an avowed externalist. Externalism will be discussed in Chapter four. For now, let us continue this investigation of global internalism by considering criticism that comes from within the internalist ranks.

More Criticism: Dancy's "Pure Cognitivism"

A somewhat similar sounding criticism of the global construal of strong internalism comes from Jonathan Dancy. Unlike Milo, Dancy is sympathetic to the cognitivist internalist approach to moral psychology. Indeed, much of his criticism of the approach exemplified by McNaughton and McDowell is that it is not a sufficiently purified form of cognitivism.[49] According to Dancy, this approach *concedes* too much to the traditional Humean account by accepting the claim "that if a cognitive state is anywhere sufficient for action, it is everywhere sufficient and so it must be impossible for a person in that state not to perform that action."[50] The concession to Humeanism that troubles Dancy is the acceptance of "the

---

[49] *Moral Reasons*, ch. 1-2. In his discussion there, Dancy offers a categorization of cognitive views that is different from the one presented here. Dancy labels approaches like McNaughton's and McDowell's, as well as any others that are inspired by the Nagelian construal of desires as consequentially ascribed states, as 'hybrid theories.'

[50] *Moral Reasons*, 53.

Humean notion of an essentially motivating state."[51] An essentially motivating state is one that motivates necessarily, under all circumstances. Such a state is 'Humean' according to Dancy since he takes the paradigm of essentially motivating states to be the conception of a conative state (paradigmatically a desire) operative in the formulation of the Humean theory of motivation. By incorporating this kind of state into their theory of motivation, McNaughton and McDowell are committed to denying the possibility of weakness and what Dancy calls the "problem of accidie," the condition of coming not to care for a while about things that one normally does care about and conceives of as being good reasons for action.[52] The person who suffers from accidie is one that we can characterize as being temporarily amoral: they are in possession of their moral judgments but are completely unmotivated by them.[53] For Dancy, the person suffering from weakness and the person suffering from accidie are part of our moral landscape and must be explained. The fact that the psychology McNaughton and McDowell employ cannot account for them renders that psychology unacceptable.

Dancy sees the problems of weakness and accidie as forming a two-pronged attack on the cognitivist internalist position. Accidie he takes to be a problem for internalism proper,

---

[51] *Moral Reasons*, 54.

[52] *Moral Reasons*, 4-5.

[53] Interestingly, Dancy follows McNaughton in thinking that a strong cognitive internalist need not worry about the amoralist proper (or the wicked agent). Both think that the claim that such agents are possible is really no more than an assertion of the externalist position and should simply be denied by the internalist. For amoralism to be problematic, Dancy contends, it would be necessary to admit that it is possible for someone to both "understand" and "accept" the claims of morality and be unmoved. But this we need not (and should not) do. All that we need say is that the amoralist is one who rejects the institution of morality altogether. Since on such a conception the agent is not accepting the claims of morality, his lack of motivation to act according to those claims does not spell trouble for the internalist doctrine. Apparently it has been lost on Dancy and McNaughton that an identical strategy is available to the externalist: the denial of the amoralist originally described amounts to nothing more than an assertion of the motivational efficacy of moral judgments. Since externalists deny that moral judgments are intrinsically motivational they should reject the denial of the amoralist. I will discuss amoralism at greater length in the next chapter. For Dancy's and McNaughton's views see *Moral Reasons*, 5-6, and *Moral Vision*, ch. 9.

which "focuses our attention on people who are not at all motivated by moral reasons which in some sense they recognize."[54] Weakness, on the other hand, raises trouble for the particular form of internalism we have been discussing in this chapter, namely the strong version that sees cognitive states as themselves motivational, that is what has come to be known as the cognitive theory of motivation. Dancy summarizes the argument against cognitivist motivation in the following way.

> Take an agent's total cognitive state and suppose that on this occasion it is sufficient for action. We must admit that the same state can be present without leading to action, because of weakness of will. But this surely contradicts our hypothesis that the state concerned was sufficient for action in the first place. We must have given an incomplete specification of that state, and since we have exhausted cognitive resources, we are left presuming that there was a non-cognitive element present as well. But to admit this is to abandon the cognitive theory altogether. The upshot of this argument is that the phenomenon of weakness of will disproves any form of cognitivism in the theory of motivation.[55]

Dancy's response to this argument is to claim that it "is unsound because it makes a basic assumption that the cognitivist does not need to share."[56] Motivational theories like McDowell's which do make the assumption of the 'Humean' thesis that motivational states are *necessarily* motivating are susceptible to this argument. By rejecting this view of motivating states, Dancy maintains, the cognitivist about motivation can adequately account for weakness of will. Furthermore, by adopting this "purified" form of cognitivism, the internalist has a solution to the problem of accidie.

Dancy's proposal is to reject the notion of necessarily motivating states for what he dubs "intrinsically motivating" states, "which can be present without motivating but which

---

[54] *Moral Reasons*, 22.

[55] *Moral Reasons*, 22.

[56] *Moral Reasons*, 22.

when they do motivate do so in their own right."[57] I see no reason why we should not accept Dancy's proposal. It would seem both sensible and desirable to allow for the possibility of states whose (primary) function is to motivate but can nevertheless exist without motivating, at least temporarily.[58] The question, however, is whether it helps the cognitivist internalist out of his troubles in explaining wrongdoing.

So how, exactly, *does* the notion of an intrinsically, but not necessarily, motivating state help the strong internalist explain weakness? Unfortunately, Dancy does not directly answer this question. He does, however, make use of this kind of state to solve the problem of accidie, which he apparently thinks is the more troubling, and claims that his answer to the problem of accidie is the same as to that of the problem of weakness.[59] That solution he gives as follows.

> What we assert is that a sate which is here sufficient for action may elsewhere not be. Where it is not sufficient, there will be an explanation for this. *And we have introduced no restriction on the sorts of explanation that we are prepared to countenance.* Sometimes the reason will be carelessness or inattention; sometimes it will be despair; sometimes it will be an excess of alcohol; sometimes it will be a neuro-physiological disorder; and sometimes it will be clinical depression. I see no difficulty in capturing the sorts of explanation felt appropriate for accidie in this sort of net. So accidie is a problem only for internalism as we originally conceived it; it can be handled by pure cognitivism without difficulty.[60]

---

[57] *Moral Reasons*, 24.

[58] Though I agree with Dancy that we should accept this modification (if this is indeed a modification) in our conception of a motivating state, it should be noted that it is one that the conativist can, and presumably should, avail himself to as well. This is so even if Dancy is right that the problematic conception of necessarily motivating states is commonly associated with conativist theories of motivation. That the traditional conativist (Humean/commonsense) theorist about motivation has expressed his view in terms of necessarily motivating desires should no more preclude him from modifying his thesis in the way suggested than the fact that the traditional cognitivist about motivation has expressed his view in terms of necessarily motivating beliefs precludes him from so modifying it. For an argument to this affect, see Steven Arkonovich, "Defending Desire: Scanlon's Anti-Humeanism," *Philosophy and Phenomenological Research*, *op. cit.*, 502-7.

[59] *Moral Reasons*, 21.

[60] *Moral Reasons*, 25-6. Emphasis added.

Whether we accept such explanations for the phenomenon of accidie will ultimately turn on whether we understand accidie to be a *moral* problem or not. We may be content to understand accidie strictly as a *psychological* failing and, if so, explaining such a failing in terms of some neuro-physiological disorder or clinical depression will not strike us as particularly problematic since psychological failings are not normally thought to be blameworthy. But it is not obvious, nor do I think it is true, that we view weakness of will, or moral weakness at any rate, in these terms. Moral weakness is a moral failing and is therefore blameworthy. By appealing to phenomena like depression, despair, and the like to explain weakness Dancy's position becomes quite problematic. Indeed, he faces difficulties that are rather similar to those facing McNaughton.

Dancy's liberalism about explanations of the failure of cognitive states to motivate opens his account up to the objection that he has given up the concept of wrongdoing. Insofar as we view wrongdoing as behavior that the wrongdoing agent is responsible for, and therefore can be held accountable and *blamed* for, we must rule out neuro-physiological disorders and clinical depression as explanations of the wrongdoer's behavior. Unless we wish to maintain the phenomenon of wrongdoing is radically confused and is almost always attributable to such psychological conditions and defects, we need to be discriminating about the kinds of explanations we allow for a moral agent's failure to be motivated to do what he believes is right. Non-moral psychological failures should not be conflated with moral ones. But this is precisely what Dancy is doing.[61]

---

[61] There is a deeper problem here for the cognitivist about moral judgment, namely *how*, in principle, to distinguish between non-moral psychological failure and moral failure. That there should be a principled way to distinguish between them would seem to follow from cognitivism about moral judgment. As we shall see, this problem will continue to plague the various versions of cognitivism we consider in subsequent chapters.

Another problem with Dancy's analysis is that it seems to overshoot its mark when applied to the case of weakness. Accidie was defined as the temporary loss of *all* motivation to do what one believes to be morally right. But in case of weakness the agent is *not* deprived of the motivation to do what is right. Instead the moral motivation is 'overcome' or 'overwhelmed' and ultimately 'defeated' by a contrary motivation. This, however, may not be a particularly troubling point for pure cognitivism. Indeed, the conativist must equally allow for motivational states that rob moral motivation of all its force (if it is to account for wickedness, indifference, and evil) and for motivational states that do not but are nonetheless capable of defeating moral motivation (in order to account for weakness). The conativist and the cognitivist about motivation are in the same boat on this point. But this directs our attention back to the real issue facing the pure cognitivist: given the conceptual resources available, how can the motivation to do wrong and the ineffectualness of the motivation to do what is right be accounted for?

Among his list of potential explanations of the failure to be morally motivated Dancy includes carelessness and inattention. As we saw in the discussion of Socrates, in order for such explanations to be explanations of *wrongdoing* the cause of the inattention must be motivational. The problem for Socrates, though, was that motivation is necessarily tied to evaluative conception, hence the inattentive agent is not so *because* he is insufficiently motivated to keep abreast of the moral details but is insufficiently motivated *because* he has an incorrect evaluative conception. But this is to give up the possibility of accounting for weakness, wickedness, and the like. Dancy is sensitive to this and he is equally aware that it appears that McDowell has fallen into this very same trap. As Dancy sees it, McDowell's claim that the virtuous agent's moral conception "silences" competing conceptions renders

weakness (and other forms of wrongdoing) impossible.[62] This is so because, as McDowell concedes, if the agent acts in ways contrary to what is morally required it can only be because he does not see that there *is* such a requirement in the situation. It is precisely to avoid this result that Dancy introduces the notion of an intrinsically—but not necessarily—motivating cognitive state. Dancy realizes that for many, if not most, of the commonly understood explanations of wrongful behavior it is essential to allow the wrongdoer to have as clear a conception of the moral terrain as the virtuous agent. But it is apparent that Dancy has missed the true nature of the problem. Carelessness and inattention imply that the agent *misses* the morally relevant features of his situation, features that the virtuous agent clearly grasps. But this implies either ignorance or negligence, not weakness or wickedness. In appealing to a psychology of besires, Dancy has put himself into a bind that seems impossible to get out of.

We can clearly see Dancy's difficulties simply by considering the very nature of his proposed remedy to the problems facing McDowell and McNaughton, namely his own "pure cognitivism," a view that only serves to recapitulate those philosopher's mistakes. For prior to Dancy's list of potential explanations of the motivational failure of intrinsically motivating states he claims that "if a consideration which succeeds in one place fails in another, there will be an explanation of why it fails. And this explanation will appeal to the content of some state which is present in the second case and not in the first."[63] But to say this is to say nothing more than what McNaughton did when he explained weakness by appeal to a competing conception that defeated the 'overall' conception. Dancy's (and McNaughton's) proposal is an advance over McDowell's in that it keeps the correct moral

---

[62] *Moral Reasons*, 59, n.15.

[63] *Moral Reason*, 24.

conception present while according to McDowell's 'silencing' picture it is replaced but, like McNaughton, Dancy is faced with the problem of explaining *how* the second conception can render the first motivationally moot. Given that motivation is tied to content, the only thing available to him is to claim that the second conception—the conception on which the agent acted—was more compelling than the first. But in saying this, Dancy, just like McNaughton, has lost wrongdoing from the inside: an agent who acts on what is to him the best conception of the situation is not one who is acting from moral weakness. Dancy's advocacy of pure cognitivism and the introduction of intrinsically but contingently motivating cognitive states has not really moved the debate at all.

Neo-Aristotelianism

It is time we asked whether there is in fact any way out for the strong internalist. Dancy objects that McDowell's cognitivism about motivation is actually a hybrid of cognitivism and conativism: moral motivation is cognitive and non-moral motivation is conative. Though I have been treating McDowell's theory of motivation as a version of global strong cognitivist internalism, it must be admitted that McDowell is less than consistent on this point. McDowell seems as sensitive as his critics to the danger of denying explanations of wrongdoing such as weakness. And he is not without a suggestion as to how to alleviate the difficulty. Yet his way out of the problem must, in fact, be seen precisely as Dancy claims, namely as advocating a mixed or hybrid form of cognitivism. While arguing for the view that the virtuous agent's moral conception silences competing conceptions, McDowell offers the following.

> It is evident that this view of virtue makes incontinence problematic. The weak incontinent person must conceive the circumstances of his action in a way which, in some sense, matches the way a virtuous person would conceive them, since he knows that he is not acting as virtue demands. But the virtuous person conceives

> the relevant sorts of situation in such a way that considerations which would otherwise be reasons for acting differently are silenced by the recognized requirement. If the incontinent person has such a conception, how can those considerations make themselves heard by his will, as they do? Obviously continence poses a parallel difficulty. The way out is to attenuate the degree to which the continent or incontinent person's conception of the situation matches that of a virtuous person. Their inclinations are aroused, as the virtuous person's are not, by their awareness of competing attractions: *a lively desire clouds or blurs the focus of their attention on "the noble".*[64]

But just how does McDowell conceive of "a lively desire"? Is it really a besire, an evaluative belief? Or is it a "Humean" desire, a second, distinct form of motivation available to the agent? Dancy, as we have seen, claims that McDowell would have it be the latter. If this is so, if McDowell is here advocating a form of mixed cognitive internalism, then he has forsaken the simplistic rational psychology of Socrates for the more complex picture offered by Aristotle. That McDowell does have this strategy in mind would seem to be evident from the following passage from "Virtue and Reason."

> If we are to retain the identification of virtue with knowledge, then, by contraposition, we are committed to denying that a virtuous person's perception of a situation can be precisely matched by someone who, in that situation, acts otherwise than virtuously. Socrates seems to have supposed that the only way to embrace this commitment is in terms of ignorance, so that, paradoxically, failure to act as a virtuous person would cannot be voluntary, at least under that description. But there is a less extreme possibility, sketched by Aristotle. This is to allow that someone who fails to act virtuously may, in a way, perceive what a virtuous person would, so that his failure to do the right thing is not inadvertent; but to insist that his failure occurs only because his appreciation of what he perceives is clouded, or unfocused, by the impact of a desire to do otherwise.[65]

McDowell is here following Aristotle's criticism of the Socratic treatment of wrongdoing, offered in *NE* VII. But though this may appear to be a welcome move, it really does not constitute an advance over the Socratic position. Aristotle's criticism and reformulation of the Socratic thesis in merely cosmetic.

---

[64] "Are Moral Requirements Hypothetical Imperatives?", 28. Emphasis added.

[65] "Virtue and Reason," 145.

Aristotle's psychology makes room for both "rational" and "non-rational" forms of desire.[66] Both kinds are motivationally efficacious, and both kinds involve evaluative conceptualization. The difference between the two resides "solely in whether or not the source of these thoughts lies in *reasons* one thinks there are for having them."[67] That is to say, rational motivation is motivation that is based upon one's reasoning into the truth of evaluative matters, whereas non-rational motivation, though involving evaluation, is not rooted in reasoning about that evaluation. The process of becoming virtuous, according to Aristotle, involves the habituated persuasion of non-rational motivation to conform to reasoned motivation. That these two forms of motivation can come apart is evidenced, Aristotle argues, by existence of the weak and the strong willed person.[68]

Aristotle's discussion of wrongdoing begins with a statement of the Socratic position and a seemingly unvarnished assessment of its inadequacy. "For Socrates was entirely opposed to the view in question, holding that there is no such thing as incontinence; no one, he said, when he judges acts against what he judges best—people act so only by reason of ignorance. *Now this view plainly contradicts the observed facts*...."[69] But those who turn to Aristotle looking for an ultimately more satisfying picture of wrongdoing are destined to come away unsatisfied, for Aristotle is as beholden as Socrates to the intrinsic motivational efficacy of evaluative judgments, especially judgments grounded in rational deliberation. Since it is precisely such evaluative judgments that Aristotle takes to be essential to moral evaluation, he is as committed to conflating wrongdoing and ignorance as was Socrates and his

---

[66] See John M. Cooper, "Some Remarks on Aristotle's Moral Psychology," in his *Reason and Emotion: Essays on Ancient Moral Psychology and Ethical Theory* (Princeton: Princeton University Press, 1999), 237-251.

[67] "Aristotle's Moral Psychology," 244.

[68] *NE* I 13 1102b15-25.

[69] *NE* VII 2 1145b25-27. Emphasis added.

contemporary descendants. We can see plainly the grip this rationalist framework has on Aristotle simply by continuing the passage quoted above. For though explaining wrongdoing in terms of ignorance "plainly contradicts the observed facts...we must inquire about what happens to such a man; if he acts by reason of ignorance, what is the manner of his ignorance?" That ignorance is the proper way to explain wrongdoing was something that Aristotle never really doubted. But there was something unsettling about the rather stark presentation of this thesis in *Protagoras*. Aristotle's aim in his discussion of wrongdoing is to render the ignorance thesis more palatable. This he attempts to do through a distinction between *types* of ignorance.

Although the Socratic discussion of *Protagoras* was aimed ostensibly at discrediting explanations of wrongdoing that appeal to weakness, the moral psychology presented there, as we have seen, serves to discredit *all* commonsense moral explanations: ignorance is all that is possible. I think it likely that it was this monochromatic rendering of the moral sphere that offended Aristotle most. He was far too keen an observer—and respectful—of the human condition and of the common practices associated with it to leave the Socratic picture unchallenged.

For Aristotle there is a clear distinction between weakness and wickedness. The wicked, Aristotle claimed, act as they *choose* whereas the weak act *against* choice, hence not willingly.[70] This distinction was an essential one for Aristotle, without which only a muddled picture of the ways of wrongdoing could result. True, weakness and wickedness are also distinguished by post-action psychological differences, for the weak are prone to regret their behavior, unlike the wicked.[71] But this is not enough. Nothing really prevents the author of a piece of

---

[70] *NE* VII 8 1151a6-7. See also J. J. Walsh, *Aristotle's Conception of Moral Weakness*, 116-7.

[71] *NE* VII 8 1150b29-30.

wicked behavior from being repentant later. It is a contingent matter whether they feel so or not. The weak, however, if they really are so, would necessarily seem repentant. But this fact is appropriately traced to their initial unwillingness to act in the way they do. Nevertheless, this is an apparently unattractive position. Contemporary discussions of weakness rarely allow for it.

McNaughton provides a particularly representative example of what I would call the contemporary sin of conflation. Attempting to set "a condition for an adequate account of weakness of will," McNaughton states that "what is needed, if we are to retain our sense of the weak-willed man as an agent, is an account which shows how the action which he judged worse could appear sufficiently attractive to the agent for him to choose it, contrary to his own best judgment."[72] The contemporary literature on the subject of weakness is rife with similar conceptions of weakness that have the weak-willed agent *choosing* the worse course, or acting 'intentionally' against his better judgment.[73] No explanation of the phenomenon of weakness, I would maintain, could possibly prove satisfactory if it *conceives* the phenomenon in this way. For this conception fails to distinguish this type of wrongdoing from wickedness and evil, and these *are* so distinguished in our moral practice. If an agent performs some action and she chose or intentionally performed it, in what sense could her will be said to be *weak*? In what sense could she be said to be *incontinent* or lacking control?

---

[72] *Moral Vision*, 123.

[73] The *locus classicus* for this way of putting the phenomenon of weakness is Donald Davidson's "How is Weakness of the Will Possible?" There Davidson defines incontinence in the following way.

> D. In doing x an agent acts incontinently if and only if: (a) the agent does x intentionally; (b) the agent believes there is an alternative action y open to him; and (c) the agent judges that, all things considered, it would be better to do y than to do x. (22)

It is hard to imagine how anyone could think that such a definition could adequately pick out the class of incontinent actions. For putting aside the question of whether a weak-willed agent acts intentionally, the definition would seem applicable both to agents who simply do not care that x is a worse alternative to y and those who would choose x *because* it is a worse alternative to y. That is, Davidson's definition is so coarse as to include *both* wicked and amoral agents whether or not it includes weak ones.

An agent that chooses or intentionally performs some action does precisely as she chose or intended to do. If the action she performed was not what she knew to be the morally correct one then she is, perhaps, wicked, indifferent, or evil, but undoubtedly not weak.

The distinction between the unwilling weak agent and the willing evil agent is rooted in our practice and is therefore a datum for theory to explain. A theory that either ignores or misconstrues that datum is therefore problematic in that respect. Taking motivational states to be evaluative belief as, for example, McNaughton does, is a particularly insidious way of being forced to misconstrue the data of wrongdoing. Choice, what Aristotle called "deliberate desire for things in our own power,"[74] essentially involves both cognitive and conative capacities. We need to be able to represent various alternatives and assess them in some way, an assessment that will be the basis for selecting among them. Such a selection will, presumably, involve motivation to act on that selection. To have chosen a course of action in this way is to have an *intention* to so act. What is wanted to capture the conception of weakness in our moral practice is the further ability to have *other* motivation, sufficiently strong, to cause the performance of an unchosen alternative. This ability would not only capture weakness but would also distinguish it from wickedness where the act performed *is* chosen. This, however, is impossible if to be motivated towards an action sufficiently so as to perform it just is to assess it as the best alternative, that is, to choose it. This problem is endemic to the strong cognitivist internalist position.

It is important to see that to ascribe an action to an agent that she did *not* choose is not to deny that her action is *free*. Many of our actions are not chosen. We do not always 'reason' about what to do, that is, consider various alternatives in an attempt to decide the best course to achieve some end. Rather, a great many of our actions are the result of habits

---

[74] *NE* III 3 1113a10-13.

and dispositions, issuing from ingrained motivational sets while a great many others proceed with immediacy from occurrent emotional states and desires.[75] When agents act from such motivations they do not deliberate, they do not act from an expressly formed *intention*—to claim that they do serves only to trivialize that notion—but their actions are free nonetheless. We hold them as responsible as they hold themselves.[76] Their motivation, obviously strong enough to usurp the motivation to do what they intend, is not irresistible: the weak are not *compelled.* Strong motivations can be overcome, sometimes by taking them head on, other times rather indirectly.[77] Either way, to deny that the weak choose their behavior or acted intentionally is not to deny their agency and responsibility.

But back to Aristotle. Though we may applaud his commitment to the distinction between weakness and wickedness, between the willing and the unwilling, we must inquire whether, given his own adherence to strong cognitive internalism, he can maintain the distinction.[78] The answer is clearly that he cannot. Aristotle's way around the conflation of weakness and wickedness that he found in Socrates was both simple and clever: distinguish

---

[75] We find this distinction nicely drawn in Hume's discussion of 'calm' and 'violent' passions (*Treatise of Human Nature* II. 3. iii). Calm passions are those that cause no sensible feeling and are known more by their effects, whereas violent passions are strongly felt. But, more to the present point, Hume associates calm passions with long-term interests and violent passions with what is more immediately before the mind. To regularly act on calm passions, that is, to have "strength of mind,"agents must consider how acting on their various motivations would effect their interests. Those who do not engage in such reasoning; those who simply act on their strongest passion of the moment without reflection on the consequences, exhibit "weakness of mind."

[76] Consider the discussion of freedom and responsibility in P.F. Strawson's "Freedom and Resentment", *Proceedings of the British Academy* 48 (1962): 1-25.

[77] I have in mind here something like the notions of 'binding' and 'side bets' discussed by Jon Estler in his *Ulysses and the Sirens: Studies in Rationality and Irrationality* (Cambridge: Cambridge University Press, 1979): ch. 2.

[78] It must be admitted that Aristotle's commitment to internalism is not universally agreed upon. Sarah Broadie, for instance, takes McDowell to task on just this point. (*Ethics with Aristotle, op. cit.*, 283-7.) Broadie argues that if Aristotle thought that moral judgment, or reasoned choice, were to be essentially practical then he would not be able to say what the virtuous, the continent, and the incontinent have in common, namely a shared belief in what is right. This is of course true, but it is not clear who deserves the praise for this insight, Aristotle or Broadie. That Aristotle thinks incontinence can be explained does not thereby show he is not an internalist, as the evidence of this chapter should surely attest. Nevertheless, I will discuss Aristotle here as

between types of ignorance and thereby distinguish between types of wrongdoing. Distinguishing between types of ignorance requires a related distinction between types of knowledge, and the requisite distinction is provided by Aristotle in the form of the distinction between universal and particular facts. Theoretical knowledge concerns the comprehension or understanding of universal facts, facts that in the moral sphere would include such items as 'courageous acts are virtuous' and 'promise keeping is right.' An agent's ignorance of such universal facts might manifest itself in the adoption of false principles, such as 'breaking promises is wise policy.' For such ignorance, Aristotle claims, a person is rightfully blamed.[79] A person in possession of false principles would be expected—given the intrinsically motivational nature of such beliefs—to abide by them. It is precisely such a person, a person who intended to perform an act of promise breaking and felt satisfied upon doing so, that Aristotle denominates as wicked.

We shall return to this analysis of wickedness presently, but the point I wish to stress now is how this analysis differs from that of the morally weak. Weak agents, as we have seen, do not intentionally aim at promise breaking or some other mistaken moral principle, thinking it true. On the contrary, such agents firmly believe, perhaps even know, that promise keeping is right and they deliberately desire, that is to say they choose, to act accordingly. Yet they don't. They break their promises just as the wicked do.[80] For Aristotle, the weak are acting ignorantly, no less so than the wicked are, but there is a difference in the nature of their ignorance. The weak possess, *ex hypothesi*, the correct moral principles, so they do not suffer from ignorance of universal facts. But such facts do not

---

though he were an internalist. If this is false, then it will be the discussion presented in chapter four that will be more apt.

[79] *NE* III 1 1110b27-35.

[80] "Incontinent people are not criminal, but they will do criminal acts." *NE* VII 8 1151a10-11.

exhaust what one can be ignorant of. There are also *particular* facts, facts such as '*this* is a case of promise keeping,' or '*this* is a case of promise breaking.' It is facts such a these, that are *perceived*, that weak agents are ignorant of, usually due to some excessive emotion or appetite that leaves their perception clouded. They *have* the knowledge that the virtuous (and the strong) have, namely that actions of such-and-such a type are not to be done but, unlike the virtuous and the strong, the weak do not *use* this knowledge because of their confused perception of particular facts.[81] Given this way of picturing the weak agent, the pain and remorse he exhibits upon the realization of his misperceptions becomes not only explicable but expected. Hence Aristotle's distinction between kinds of ignorance enables him to at once stay within the Socratic tradition of viewing virtue as moral knowledge, yet remain more faithful to the facts of our moral experience: an agent can have an awareness of what is virtuous and still be moved by the motivation to do otherwise.

Has Aristotle offered a way of understanding and explaining wrongdoing that the contemporary strong internalists like McNaughton, McDowell, Dancy and others can adopt? I don't think so. The problems endemic to his account of the psychology of moral judgments remain. We need only scratch the surface of the Aristotelian approach to see that this is so. Aristotle repeatedly emphasizes that the morally weak act not with choice but against it. His emphasis on this is appropriate since that is precisely how it seems to the weak agent. The phenomenology of weakness has the weak agent endorsing or choosing one path only to suffer the indignity of feeling himself pulled towards another. But by claiming that the weak act as they do from a failure of perceptual knowledge this phenomenology becomes inexplicable. According to Aristotle's analysis, when an agent fails to see that *this* act would be an instance of promise breaking, they instead see it as an

---

[81] *NE* VII 31146b31-35, 1147a4-10.

instance of, say, doing something enjoyable with a friend, something which they take to be good in itself. This is supposed to be due to their strong desire to do something enjoyable with a friend or because of their distaste for keeping an odious promise. However, if their strong appetites or feelings lead them to see their action in only one way, a way that presents it as good, in what sense are such people pulled towards doing something they believe to be wrong? No doubt we can say that they are, confidently, from the third-person perspective, but this isn't really what we are after in an explanation of wrongdoing that appeals to weakness.[82] We want an explanation that captures things from the *inside*. But on Aristotle's account we don't get this. Viewed from the agent's perspective, *what it is like* to act weakly is really no different than what it is like to act wickedly (or virtuously, for that matter). Both the weak and the wicked act as they do believing that their particular act is an instance of a general principle or type which they endorse. To claim that a conative state leads to the "clouded perception" of what is appropriate, as Aristotle and McDowell wish to maintain, results in a denial of the phenomenology of weakness. The mixed version of strong internalism fares no better with this type of wrongdoing than the global version does.

And what of wickedness? The limitations of the Aristotelian notion of wickedness has been nicely adumbrated by Milo in his book *Immorality*. As will be recalled from the sketch of types of wrongdoing Milo finds in our commonsense moral practice (Chapter one), we can distinguish what we might call *perverse* wickedness from *preferential* wickedness. Perverse wickedness is precisely what Aristotle is offering us: an agent who is intentionally and willfully doing what is wrong (hence his behavior is *wicked*), but who nonetheless believes his action to be right (because he is ignorant). Whether we allow for such an explanation of

[82] At least it is not the only thing we want.

wrongdoing it should be clear by this point that it cannot be the whole story.[83] Surely we must allow for those who willfully do what is wrong while *believing* it to be wrong. And this is not simply because we think that there are such people but because the possibility of such a psychological explanation is integral to the commonsense picture or wrongdoing that includes differences in *degree* of moral turpitude. Milo captures this point clearly in the following.

> It is essential to our notion of wickedness as a more evil or blameworthy kind of wrongdoing that the agent of a wicked act does not desire to avoid moral wrongdoing, since this would mean that he has the same redeeming quality as the agent of the morally weak act. For this reason, it seems better to conceive of the agent of a wicked act as willingly doing something that he himself believes to be wrong.[84]

Aristotle does not offer a fundamentally different picture of the wrongdoing agent than Socrates, and McDowell cannot appeal to his method to escape from the insuperable difficulties their shared approach faces. A mixed cognitivist internalism is no better off than a global, or pure, version.

If the foregoing is correct, we have identified a significant fact about the psychology implicit in our everyday moral practice: moral judgments are not themselves motivational.[85] To claim that they are, to say that moral judgments are properly understood as something along the lines of a 'besire,' is to preclude a coherent understanding of *how* an agent can do wrong. This is because wrongdoing is a moral category that presupposes that its source is in motivation, not in judgment. That is to say that the concept of wrongdoing is incompatible with a conception of moral judgment that *both* aims at truth *and* intrinsically moves an agent

---

[83] Nothing, so far as I can tell, should prevent us from accepting that some people do in fact think that what they are doing is right when they are doing wrong and hence act willfully. But Aristotle's claim that they *ought* to know better notwithstanding, I see no reason why we should think of such people as *wicked*.

[84] *Immorality*, 218.

[85] At least insofar as we think of such judgments as beliefs.

to act. A morality without wrongdoing is no morality at all. Unless we are content with moral skepticism, we ought to reject the rationalistic psychology of strong internalism.

3

The Good, the Bad, and the Indifferent

> M. de Roannez said: "Reasons come to me afterwards, but at first a thing pleases or shocks me without my knowing the reason, and yet it shocks me for that reason which I only discover afterwards." But I believe, not that it shocked him for the reasons which were found afterwards, but that these reasons were only found because it shocks him.
>
> The heart has reasons, which reason does not know.
>
> —Blaise Pascal[1]

In the previous chapter we argued that the metaethical position we have called 'strong cognitive internalism' is inadequate to the demand that it allow for and plausibly explain wrongdoing. It was argued in the first chapter that meeting this demand is something that a philosophic theory of ethics must do, for wrongdoing is a crucial aspect of our commonsense moral practice, both influencing the form of that practice and providing practical significance to theory construction. In attempting to construe moral judgments as intrinsically motivational, strong cognitive internalism requires that we conceive wrongdoing in radically unfamiliar ways. The upshot is that strong cognitive internalism renders impossible the very conception of wrongdoing that is among the initial prompts for theory construction. In the present chapter I will provide an analysis of a position we will call 'weak cognitive internalism.' This position attempts to keep the attractive features of both cognitivism and internalism while providing a more intuitively satisfying account of wrongdoing. The argument here will be quite the opposite: no such account can be made to work.

---

[1] Blaise Pascal, *Pensées.* (1670). Translated by William Finlayson Trotter. Reprinted in Monroe C. Beardsley, ed., *The European Philosophers: From Descartes to Nietzsche* (New York: The Modern Library, 1960): 123-4.

Foreground

Hume famously claimed that it is not "contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than the latter."[2] For Hume this claim about the flexibility of human psychology followed directly from the claim that reason—including the paradigmatic psychological state associated with that capacity, belief—and passion are essentially distinct.[3] If one takes reason to be essentially representational and motivationally impotent, and passion to be motivationally efficacious and non-representational, then no conclusions can be drawn about the relation between the two *a priori*. Coming to believe that a certain course of action will be more beneficial to oneself than another or would be morally right, in no way guarantees how one will be motivationally oriented towards that action. Indeed, a person may have such evaluative beliefs and not be motivated at all by them.

One might argue that Hume's claim only follows if we accept a purely instrumental, as opposed to substantive, conception of rationality. But there is, I believe, a non-question-begging sense of 'rational' on which we can see Hume to be correct. If so, an important and influential view in moral psychology is false. Consider the following interpretation of Hume's claim. It is not that it is *never* contrary to reason, or irrational, to be insufficiently, if at all, moved by considerations of one' own or the moral good. Rather, it is Hume's point that irrationality cannot be *essential* to such a motivational condition, not if such a motivational condition is to be subject to evaluative criticism. A person who would scratch her finger rather than avert the destruction of the whole world faces our unrestrained

---

[2] David Hume, *A Treatise on Human Nature*, *op. cit.*, II. iii. 3, 416.

[3] Given that the most natural construal of belief is that of an *attitude* towards a proposition, the question of whether belief *should* be thought of as the paradigmatic state associated with reason, or cognition, is a question that needs to be addressed but I will not do so here.

condemnation. To want the end of all we know or to show no care for its passing is to be profoundly at odds with one's fellows. This is an unacceptable practical stance. But consider closely the posture of the opposition and one finds not simply the counsel of prudence but the indignation of the affronted. Such a person faces not only our sanction but our resentment. Such motivation stands *morally* opposed. Our resentment, however, is susceptible to becoming the object of further resentment if it is not directed at the appropriate target. That is, our resentment and indignation must be *fitting*—those to whom we react in this way must satisfy the "constraints on moral address."[4] Our reaction to such motivation *is* our holding the person who is so motivated morally *responsible*. Yet, if a person who cares not for the passing of the world is *not* responsible for her indifference then our resentment is misplaced and offensive in its own right. Hume's point is that if such malice or depraved indifference *ever* warrants our indignation—if it is ever blameworthy—then it must be something that a person is *capable* of being responsible for. It must be possible that a rational person, one in full command of her powers of contemplation, reflection, calculation and choice can nevertheless be left cold by the suffering of others (even that of herself). But then such motivation cannot be essentially irrational. If it were, if such indifference were the result of some rational incapacity, it could not be morally wrong.

Hume's point is by no means restricted to the realm of wrongdoing. To sacrifice oneself—to choose one's "total ruin"—for the sake of another, even a stranger, is often the most noble and virtuous of acts. People who so act show no greater love and deserve no greater praise. But to be deserving of our praise such people must be responsible for their

---

[4] The phrase is Gary Watson's, from his "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme," from Ferdinand David Shoeman, ed., *Responsibility, Character, and the Emotions* (Cambridge: Cambridge University Press, 1987): 256-286. It will not go unnoticed that the conception of moral responsibility that will be employed here is deeply indebted to the ideas of Peter Strawson. See his "Freedom and Resentment," *Proceedings of the British Academy*, Vol. 48 (1962).

extraordinary self-sacrifice. It must be possible for people to freely choose such a course upon rational reflection. It cannot be the mark of irrationality to put the good of others (or the moral good) before oneself. If it were, the scope of moral evaluation—both positive and negative—would be diminished beyond recognition. Being capable of rational reflection, being able to appreciate the significance (particularly the moral significance) of one's situation and to act in light of that appreciation is what qualifies one for moral evaluation, what makes one an appropriate object of moral address. *How* one is morally addressed—how others *react*—depends on how one is affected, or changed, by what is taken to be significant. It must be possible then for a rational person to be affected in ways that merit both praise and blame for moral intercourse to make sense, not to mention be justified. Evaluable attitudes cannot be contrary to reason. Indeed, it is the fact that they are eminently reasonable that renders them evaluable.

As we have noted, Hume took this situation to follow from (or, perhaps, to demand) a psychic division of labor. People can be free to be moved or unmoved by the thought of the flourishing or suffering of themselves or others only if conative/affective capacities are functionally distinct from representational, or cognitive capacities. Belief—even evaluative belief—cannot entail *motivation*, let alone of a particular kind. To think otherwise is to misconstrue the nature of the moral assessment of agents, their motivations, and their actions. This means that we must follow Hume and reject the so-called 'cognitive theory of motivation' for a more classical, or 'conative' (often called, unfortunately, 'Humean') account. Indeed, our commonsense moral practice demands such a theory for that practice is firmly embedded within our commonsense, belief/desire psychology, a psychological framework that itself presupposes a conative account of motivation.

In the previous chapter we argued in this Humean manner, focusing on the difficulty cognitivism about motivation has with allowing for and explaining wrongdoing. By construing cognitive states such as evaluative belief as intrinsically motivational, 'strong cognitive internalism,' as we labeled the position, denied itself the resources to provide intuitively satisfying construals of such commonsense phenomena as moral weakness, wickedness, indifference, and negligence. All of these phenomena—indeed, all blameworthy wrongdoing—presuppose the possibility of divergence between moral belief and motivation. The inability to allow for such divergence is the price paid for the adoption of an overly rationalist psychology, one that conflates cognitive and conative capacities. Strong cognitive internalism is forced to construe moral failure in terms of ignorance just as Socrates, the ultimate rationalist, was forced to do. But this is unacceptable. Ignorance is not an ethical category: failure to know what is of moral significance involves no want of principle.[5] Cognitivism about motivation is bad psychology and as a *moral* psychology it is self-defeating.

Our moral practice, embedded as it is in our commonsense belief/desire psychology, demands that we respect the "dogma of philosophical psychology"[6] that is the psychic division of labor. Acceptance of the division serves as the backdrop for the central debate in contemporary metaethics. For being conativist about motivation does not of itself constrain one's views concerning the psychology of moral judgment: one is free to understand such judgments in either cognitivist or conativist (expressivist, or projectivist) fashion. Which way one goes on this question can be influenced by one's views about the

---

[5] Failure to have morally relevant information is blameworthy if that ignorance is due to the carelessness, inattention, or indifference of the agent herself. But this is negligence, which requires that an agent believe she ought to have certain information but is insufficiently motivated, if at all, to acquire it. If moral beliefs are intrinsically motivational negligence, too, is psychologically problematic, if not impossible.

[6] Michael Smith, "The Humean Theory of Motivation," *op. cit.*

'practicality' of value judgments in general and moral judgments in particular. Conativists about the psychology of value follow Hume in insisting on the motivational efficacy of value judgments. Since the psychic division of labor allows only conative states to be motivational they conclude that such judgments must be expressions, or functions of, conative states or some confluence of such states. Evaluative belief, on this picture, is comprehensible only against a profile of attitudes and concerns. For the conativist about moral judgment, such judgment and motivation are conceptually wed.[7]

Others, however, reject the claim that value judgments are essentially practical, justifying not even the presumption of a guarantee of motivational influence. For them, the consequences of the psychic division of labor are unambiguous: judgments of value are representational—they aim to depict an evaluative reality—while motivational states such as desires, attitudes, concerns, and emotions, are essentially distinct and only contingently accompany (if at all) such judgments.[8] These philosophers combine cognitivism about the judgment of value with conativism about motivation. Since they, following Hume, see no essential connection between representation and motivation, they (perhaps unlike Hume) hold that it is quite possible to make value judgments (moral or otherwise) without being

---

[7] Whether such a linkage undermines the plausibility of a conativist construal of moral judgments, precluding a satisfactory explanation of wrongdoing, will not be considered here. It should be noted, however, that the situation for the conativist is not symmetrical with that of the strong cognitive internalist. Tying evaluative representation to motivation is very different from tying motivation to evaluative representation. For the latter variation in motivation is only possible with variation in representation. This is not so for the former. Nevertheless, that the conativist position is in fact undermined is the charge its rivals, specifically cognitive externalism. For some criticisms of conativist explanations of wrongdoing from the externalist perspective, see David Brink's *Moral Realism and the Foundations of Ethics*, *op. cit.* 83-6; Brink's "Moral Motivation," *op. cit.*; Ronald Milo's "Virtue, Knowledge, and Wickedness," *op. cit.*; and Nicholas Sturgeon's "What Difference Does it Make Whether Moral Realism is True?," in N. Gillespie, ed., *Moral Realism: Proceedings of the 1985 Spindel Conference. The Southern Journal of Philosophy, Supplement* 24, 115-41.

[8] Whether the evaluative reality is coextensive with 'natural' reality is an open question as far as this paper is concerned. Nothing of what follows turns on answering (or not answering) that question.

motivated by them. This position is known as 'cognitive externalism' ('externalism' hereafter).

If we could trust Hume's view of the logic of the matter, conativism and externalism would exhaust the theoretical options on the issue of evaluative judgment. If conativism is true we can conclude that value is essentially practical (motivationally influential) in nature but, for that very reason, it is to some degree subject to the variability of motivation. Alternatively, if cognitivism is true we can conclude that value is an objective feature of reality (in which case our judgments of value have truth conditions) but, for that very reason, such judgments are not necessarily motivationally significant. But not all do trust Hume on this matter. Some philosophers feel that we are being offered a false dichotomy. Why must the objectivity of value be sacrificed to secure practicality, or practicality for objectivity? Some philosophers suggest, *pace* both conativists and externalists, that one can be conativist about motivation, cognitivist about evaluative judgment, *and* have license to determine *a priori* the nature of a person's motivational orientation simply from knowing the content of her evaluative judgments. Hence knowing that a person judges that it is (morally) wrong to kill another mother's children is all that one needs to know in order to be entitled to conclude that that person is (conatively) *against* such action. Moral realists, then, need not deny the practicality of moral judgment. Holders of this view are 'weak cognitive internalists.'[9]

---

[9] A note on usage. There are a number of different ways of construing the strong/weak distinction for internalism. For instance, some philosophers, most notably David Brink (*Moral Realism and the Foundations of Ethics*, *op. cit.*, 41-2), distinguish between internalists who claim that moral judgments provide *sufficient* motivation for appropriate action and internalists who claim only that such judgments provide *some* motivation for such action. Others, most notably Michael Smith (*The Moral Problem*, *op. cit.*, 61), distinguish between internalists who claim that moral judgments motivate *simpliciter* and internalists who claim that the connection between moral judgment and motivation, though necessary, is *defeasible*. As I am employing them, these terms capture the distinction between those who, denying the psychic division of labor, claim that moral judgments are *intrinsically* motivating and those who, accepting the psychic division of labor, claim that moral judgments are *relationally* motivating. This distinction is, I believe, more fundamental than Smith's, which is, to some

Why is this internalism weak? Strong internalism, as we have seen, involves the rejection of psychic division of labor and takes cognitive states, such as evaluative beliefs, to be intrinsically motivational. On this view judgments of value motivate of *metaphysical* necessity. However, this has the result of putting culpable wrongdoing beyond conceptual reach. As Michael Smith, the most prominent contemporary weak internalist, puts it, strong internalism

> commits us to denying that, for example, weakness of will and the like may defeat an agent's moral motivations while leaving her appreciation for her moral reasons intact. And for this reason it is, I think, a manifestly implausible claim as well.[10]

It is the weak internalist's aim to avoid such implausibility. They attempt to do so by construing the connection between value (in particular, moral) judgment and the will that is the hallmark of internalism as a *defeasible* one. That is, it is a connection prone to breakdown in certain specifiable circumstances. As to the nature of those circumstances we are given some idea by the very definition of the position. Offering what he labels "Weak Internalism," Smith defines the position thusly:

> *Weak Internalism:* If an agent judges it right to ϕ in certain circumstances C, then she is motivated to ϕ in C, at least absent weakness of will and the like.[11]

---

degree, consequent on it. It is of course possible for someone who accepts the psychic division of labor to maintain that the link between evaluative belief and motivation is *not* defeasible, but it is not possible for the cognitivist about motivation to claim that it is. Jonathan Dancy's proposal that evaluative beliefs are intrinsically motivating but not necessarily can be seen as an attempt to reject the psychic division of labor while securing the defeasibility of the connection between evaluative judgment and motivation that is the hallmark of the weak internalist position. If the arguments of the previous chapter are correct however, no such combination is possible. See Dancy's *Moral Reasons*, *op. cit.* ch.s 2-3.

[10] Michael Smith, *The Moral Problem*, *op. cit.* 61.

[11] Michael Smith, "The Argument for Internalism: Reply to Miller," *Analysis*, 56.3 (1996): 175-184, 175. Since I will be making extensive reference to this work and others by Smith, especially *The Moral Problem*, and "In Defense of *The Moral Problem*: A Reply to Brink, Copp, and Sayre-McCord," Ethics Vol. 108 No. 1 (1997): 84-119, I will make future references parenthetically in the text, using "RM", "MP", and "D" for these three works, respectively. Incidentally, the motivation an agent is 'required' to have with respect to her moral judgments (and judgments of value in general) is not to be thought of as necessarily 'overriding'. Hence being appropriately motivated to do what one judges right is no guarantee that one will do what one judges right. This qualification should be assumed in all subsequent formulations of putative requirements on such judgments and their relation to motivation.

This form of internalism is weak, Smith claims, because "those who accept it can agree that agents who judge it right to act in a certain way may not be motivated to act that way at all if they are weak-willed" (RM, 175). If, as would seem reasonable, other forms of wrongdoing are to be included in the "and the like" of the definition, weak internalism can be seen as an explicit attempt to allow for the possibility of culpable wrongdoing within the cognitive internalist paradigm.

It might be reasonably asked, however, why we should expect the internalist connection between moral judgment and the will to break down at all, let alone in such particular (and, theoretically speaking, rather fortunate) circumstances. The answer is supposed to lie in the nature of the connection itself. As we have noted, according to strong internalism, the motivational influence of moral judgment is metaphysically necessary since such judgments are, by their very nature, motivational. But this is not a possibility on the weak internalist hypothesis, which accepts the psychic division of labor. For the weak cognitive internalist, moral judgments and motivational states are distinct existences. Since this distinction suggests that one can make a moral judgment without the accompaniment of a motivational state we have reason to think that any connection between such states could not be anything but defeasible. Of course, there is nothing in this that a cognitive externalist could not (and does not) claim. What distinguishes the weak internalist from strong internalists, on the one hand, and externalists on the other, is the claim that the connection between moral judgment and motivation is purely conceptual. That is, moral judgments are accompanied by motivational states by *psychological* necessity, in the sense that an agent cannot make a genuine moral judgment without *also* having an appropriate motivational state (absent weakness of the will and the like). For the weak internalist there is a psychological link, severable though it may be, between moral judgment and motivation.

An example might be helpful. Imagine a woman who judges that it is morally wrong to kill another mother's children. The strong internalist claims that the very act of moral judgment has, essentially, a motivational aspect. Hence a woman who so judges that killing another mother's children is wrong is, necessarily, motivated in accordance with that judgment. That is, she has an aversion to killing another mother's children. The cognitive externalist, on the other hand, does not take moral judgment to have a motivational element, and we therefore cannot say anything *a priori* about the motivational orientation of the woman who makes such a judgment. In contrast to both of these views, the weak cognitive internalist claims that though the moral judgment that killing another mother's children is not itself motivational, we can nevertheless know *a priori* that the woman making such a judgment is appropriately motivated with respect to it *or* that she is suffering from some psychological condition, perhaps weakness of will. This is so because it is a part of the very concept of moral judgment that it be linked with the motivational capacities of the agent, absent some psychological condition that would preclude it.

It is just as important that we ask why we should expect there to *be* a connection between moral judgment and motivation, a connection that is knowable *a priori*. Why, that is, should we think, as the weak internalist does, that the 'distinct existences' of evaluative belief and motivation are necessarily related? The psychological link the weak internalist claims obtains between moral judgment and motivation stands in need of explanation if it is not to be an arbitrary and unexplained brute fact. To best understand the weak internalist's proffered explanation it is worth taking a step back from our discussion as it has been framed so far.

As we have been considering the adequacy of the different versions of cognitive internalism, we have understood them from a purely psychological perspective: whether one

is, say, a strong or weak internalist is a function of whether one takes moral judgments to be intrinsically motivational or whether one takes moral judgments to be only related to motivational states. But we might also consider such distinctions from a more traditional philosophical perspective, from that of competing metaethical positions. Viewed in this way, strong internalism is the natural moral psychology for those theorists who adopt what is often called a 'sensibility theory,' such as McDowell, Wiggins, and McNaughton. These philosophers, as we noted in the previous chapter, take the perception (the cognition) of moral facts to essentially involve what we would call a 'conative/affective component,' just as Socrates did. Without such a motivational component to cognition, true moral judgments would not be possible.[12]

The weak internalist, on the other hand, holds the relationship between moral judgment and motivation to be only relational, not metaphysical, and therefore must posit some other account for why that relation should be thought to hold. The dominant proposal comes from those who attempt to ground morality in *practical* reasoning. Following in a tradition greatly influenced by both Hobbes and Kant, such philosophers hold that what is morally right to do is a matter of what it would be most rational to do: moral facts are facts about what it is most rational to do.[13] The affinity between weak internalism in moral psychology and practical reasoning theory in morality becomes plain if one thinks that judgments of what is most rational to do necessarily produce motivation (if one is not already so motivated) to act accordingly. For if one holds internalism about judgments of practical rationality, and one holds that moral judgments are a species of rational judgments,

---

[12] See Stephen Darwall, "Reasons, Motives, and the Demands of Morality: An Introduction," in Stephen Darwall, Allan Gibbard, and Peter Railton, eds., *Moral Discourse and Practice: Some Philosophical Approaches*, *op.cit.*, pp. 305-12.

[13] See Stephen Darwall, Allan Gibbard, and Peter Railton, "Toward *Fin de siècle* Ethics: Some Trends," *The Philosophical Review*, Vol. 101, No. 1 (1992): 115-89, especially 131-7.

internalism about moral judgment seemingly drops out.[14] What explains the necessary relationship between moral judgment and motivation for the weak internalist then is the apparent truth of these two further positions: internalism about practical reason and 'practical rationalism' about morality. It is natural, then, to call those who hold these two theses 'weak internalists.'

Consider the following understanding of internalism offered by Christine Korsgaard.

> Thus, it seems to be a requirement on practical reasons, that they be capable of motivating us. This is where the difficulty arises about reasons that do not, like means/end reasons, draw on an obvious motivational source. So long as there is doubt about whether a given consideration is able to motivate a rational person, there is doubt about whether that consideration has the force of a practical *reason*. The consideration that such and such action is a means to getting what you want has a clear motivational source; so no one doubts that this is a reason. Practical-reason claims, if they are really to present us with reasons for action, must be capable of motivating rational persons. I will call this the *internalism requirement*.[15]

As Korsgaard sees it, for something to be a practical reason, a *rational* agent must be motivated by it, for an agent who is *not* so motivated by such a consideration cannot count as fully rational (if that consideration is really a reason). But this analysis of reasons (and of rational agents) is not meant to stand alone, Korsgaard claims, but can be employed for the purposes of ethical theory. As she puts it, "the force of the internalism requirement is psychological: what it does is not to refute ethical theories, but to make a psychological

---

[14] See David Brink, "Moral Motivation," *Ethics*, pp. 15-6.

[15] Christine Korsgaard, "Skepticism about Practical Reason," *Journal of Philosophy* LXXXIII (1986): 5-26, quotation from pg. 11. Emphasis in the original. It must be admitted that Korsgaard's commitment to internalism about practical reason (and, in turn, morality) is ambiguous. As the quoted passage indicates, her understanding of the 'internalism requirement' is that practical reason claims must be "capable" of motivating rational agents. In an important sense this requirement is vacuously weak: it is not obvious who, if anyone, would reject it. For instance, it is no part of motivational externalism (about morality or reasons) to claim that judgments of reason (or morality) are *incapable* of motivating rational agents, only that they do not *necessarily* motivate. Yet even in later work that is overtly critical of what she calls "dogmatic rationalism," she maintains the view that reason claims will motivate an agent "in so far as he is rational." See her "The Normativity of Instrumental Reason," in Garrett Cullity and Berys Gaut, eds., *Ethics and Practical Reason* (Oxford: Clarendon Press, 1997): 215-54. Now if moral judgments are judgments of practical reason, it then follows that agents are motivated by their moral judgments *in so far as they are rational*. And it is with this view that I intend to take issue with here.

demand of them."[16] The demand is that legitimate ethical claims be motivational for rational agents. If ethical claims are conceived as claims about practical reason then this can be assured.

Investigations into this line of thinking have become increasing popular in the philosophical literature. This has led Garrett Cullity and Berys Gaut, editors of a volume on ethics and practical reason, to claim that the idea that there is a "conceptual connection between normative reasons and motivation is common ground to contemporary theorizing about practical reason."[17] The nature of that conceptual connection they express as follows:

> A normative reason for me to ϕ must be a consideration my awareness of which would motivate me to ϕ if I were thinking about it fully rationally and with full knowledge.[18]

Clearly, if moral reasons are a species of normative reason, then we should expect (fully) rational agents to be motivated by their judgments of such reasons.

Among the primary contemporary inspirations for this line of thinking about morality and motivation is the work of Thomas Nagel.[19] The following passage from *The Possibility of Altruism* makes this plain.

> This book defends a conception of ethics, and a related conception of human nature, according to which certain important moral principles state rational conditions on desire and action which derive from a basic requirement of altruism....If the requirements of ethics are rational requirements, it follows that the motive for submitting to them must be one which it would be contrary to reason to ignore.[20]

---

[16] Ibid., pg. 23.

[17] Garret Cullity and Berys Gaut, eds., *Ethics and Practical Reason*, *op. cit.*, pg. 3.

[18] Ibid., pg. 3.

[19] As we saw in the previous chapter, Nagel has proved inspirational to strong internalists like McDowell and Dancy as well. According to Smith, Nagel is best understood as a strong internalist, not a weak one. See *The Moral Problem*, pp. 211-12 (n. 12).

[20] Thomas Nagel, *The Possibility of Altruism*, *op. cit.*, pg. 3.

It is with this view, that judgments of morality are properly seen as judgments of practical rationality, which are in turn seen as necessarily motivational, that I will take issue with here. According to such a view, failure to be motivated according to one's moral judgments—what we would pre-theoretically take to be immoral—is properly understood as a matter of being (practically) irrational. This, I will argue, is an unacceptable analysis of immorality and any moral psychological or metaethical view that entails it ought to be rejected. For the sake of simplicity it will be useful to focus the following discussion on the writings of Michael Smith, currently the most prominent defender of this position and the one who has done the most to develop it.[21]

The structure of this chapter is as follows. In the following two sections I will look more closely at the nature of the circumstances in which the psychological link between moral judgment and motivation is supposed to break down. I will argue there that such circumstances, far from providing an intuitively satisfying analysis of culpable wrongdoing—among the very circumstances in which, according to commonsense moral practice, any connection between moral judgment and motivation *does* break down—, actually have little or nothing to do with moral failure. In the final section I will consider what Michael Smith offers as the primary argument for weak internalism, an argument meant to show that weak internalism provides the only intuitively satisfying account of morally *praiseworthy* behavior. Building upon the argument of the previous section, I will argue that, far from providing such an intuitively satisfying account of such behavior, weak internalism actually serves to undermine moral success and in a manner complimentary to its misconstrual of moral failure. The conclusion to be arrived at is that weak internalism is no more successful than strong internalism at providing an intuitively satisfying moral psychology. A necessary

---

[21] Smith is, at any rate, the most unambiguous adherent of this position (see notes 15 and 19, above).

connection between moral judgment and motivation, be it metaphysical or conceptual, entails a distorted picture of our moral practice. The upshot will be that the moral realist should advocate an externalist moral psychology.

Immorality as Irrationality, Part I

Smith claims that Weak Internalism has the status of a conceptual truth (RM: 175). That is to say that it is part of our concept of moral judgment that agents who make such judgments are "motivated accordingly, at least absent weakness of will and the like" (MP: 66). Now if, as we said above we might reasonably expect, phenomena such as wickedness, indifference, and negligence were to be the content of "and the like," there would be little to quarrel with concerning this 'conceptual truth.' Indeed, we might rephrase Weak Internalism, putting it in the form of a conceptual truth, or requirement, of metaethics. Let's call this metaethical requirement MR.

> MR: If an agent judges it right to ϕ in circumstances C, then either she is motivated to ϕ in C or she is immoral (providing her failure to be motivated is not due to some psychological disorder for which she is not responsible, nor beyond her ability to overcome).[22]

MR is, of course, true but it is not a particularly interesting claim or one that needs much by way of defense. This is because MR is simply a formalization of a constraint implicit in our very practice of morally evaluating agents. This is to say that MR expresses, or reflects (a part of) the contour of our morally evaluative practice, not that it is a principle which

[22] Two points concerning this principle. First, it is not clear whether the motivation an agent must exhibit in order to be moral must be overriding or not. Weakness, *pace* Smith, does involve motivation in accordance with moral judgment, yet it is not overriding. Second, the term 'immoral' ought to be interpreted widely, such that an agent may qualify as immoral on a particular occasion, as opposed to the more narrow interpretation where that term is reserved for morally blameworthy character.

determines, or guides that practice.[23] Put another way, MR is a principle that any metaethical position must respect on pain of not being a meta*ethical* position. MR is, to that extent, uncontroversial. MR states a conceptual truth because immorality (i.e. culpable wrongdoing) *just is* inappropriate (including the absence of) motivation with respect to moral judgment.[24] If Weak Internalism is simply a less perspicuous expression of MR then it, too, is an unassailable, and uncontroversial, truth.

Weak Internalism is, however, neither unassailable nor uncontroversial and Smith does not offer it as such. For Weak Internalism is not a truth *about* metaethics but a position *within* metaethics, and a tendentious one at that. The more perspicuous formulation of Weak Internalism is found in *The Moral Problem*, labeled "the practicality requirement on moral judgment" (PR, hereafter) and worded as follows.

> PR: If an agent judges that it is right for her to ϕ in circumstances C, then either she is motivated to ϕ in C or she is practically irrational (MP: 61).[25]

Smith accompanies this definition with the following elaboration:

> In other words, agents who judge it right to act in various ways are so motivated, and necessarily so, absent the distorting influences of weakness of will and other similar forms of practical unreason on their motivations (MP:61).

And with this it becomes clear that PR (and Weak Internalism) is not an innocuous statement of the obvious but a substantive account of *why* MR is true. PR is clearly meant to serve an explanatory function; Smith is offering PR as the best explanation for the truth of MR. In order to fulfil this function PR must satisfy two conditions. The first is the

---

[23] Nor is it a principle that is justifiable independently of that practice (nor, for that matter, is it justifiable *by* that practice—it isn't justifiable at all). The principle is nothing more than a formulation *of* that practice.

[24] This isn't quite right, as we shall see in the next section. The subtlety of our evaluative practice is such that not all instances of failure of motivation with respect to moral judgment (for which the agent can be held accountable) will count as immoral. This complication can be innocently ignored for present purposes.

[25] Compare the passages from Korsgaard, Nagel, and Cullity and Gaut quoted in the previous section.

rather obvious one that PR be *consistent* with MR. In other words, if PR is to explain the truth of MR, it must be the case that PR and MR can be *both* true. If the truth of PR requires the falsity of MR then, quite obviously, PR cannot account for the truth of MR. Since MR is true, this would refute PR. The second condition PR must satisfy, though perhaps less obvious, is no less important. If PR is to serve as an *explanation* of MR's truth, not only must PR be consistent with MR, but any defense of PR must show how the truth of MR *presupposes*, or is a necessary condition for, the truth of PR. It cannot be the case that the cost of making PR and MR consistent leaves PR unable to explain MR by either having PR *presuppose* MR, or by the fact that they are essentially independent. Satisfying both of these conditions simultaneously, as we shall see, is no easy task.

The question of the plausibility of the hypothesis that there is a psychological link between moral judgment and motivation, a link that, as we might say, 'defeasibly guarantees' the appropriate motivation in the morally astute agent, can now be recast as the question of the plausibility that such a link is underwritten by the concept of practical rationality. As we have seen, anyone who thinks that moral judgment is (metaphysically) distinct from motivation yet connected to it in a principled way (i.e. weak internalists) is obligated to explain the nature of this link. The explanation that weak internalists give is embodied in PR. Having appropriately linked moral judgments and motivational states is constitutive of being *practically rational.* So long as moral agents are functioning rationally they will be appropriately motivated with respect to their moral judgments. Insofar as they are inappropriately motivated—that is, insofar as they are immoral—they are practically irrational. Whether this claim is at all plausible depends, obviously so, on precisely what practical rationality/irrationality amounts to.

Clearly what irrationality, in this context, cannot amount to is a significant deficit in the ability to reflect rationally upon one's situation, circumstances, options, consequences, and the like. This is simply the point we encountered earlier in discussing Hume's claims about what is and what is not contrary to reason. Our concept of immorality (and, of course, morality) is bound up with that of responsibility. In turn, our concept of (moral) responsibility is bound up with the ability to contemplate and reflect in ways that can influence action and choice. An agent who is irrational in the sense of having lost this ability would hardly count as morally responsible for her actions. Indeed, she may not qualify as an *agent* at all. A creature rendered incapable of such reflection would not count as an appropriate object of evaluation, nor would her behavior. If 'practically irrational,' as it appears in PR, were to refer to such a cognitive deficit then PR would, quite obviously, be *inconsistent* with MR. PR and MR could not *both* be conceptual truths according to such a reading. If, as this interpretation of PR would have it, agents only fail to be appropriately motivated with respect to moral judgments when they are unable to reflect rationally on the significance such judgments have for their circumstances and options then it is not the case that such inappropriately motivated agents are necessarily immoral. Assuming the unassailable status of MR, PR would have to be rejected as false.

The defender of PR might plausibly object that inability to reflect rationally significantly undermines the ability to *form* moral judgments, hence construing practical irrationality as a *cognitive* failing badly misconstrues PR, rendering it virtually incoherent. PR, after all, is meant as a conceptual claim about the connection between moral judgment and motivation, between cognition of value and conation. That the cognitive end of this relation has gone off without a hitch is meant to be a given: the agent has made a moral judgment. The claim is that the appropriate motivational orientation will follow suit unless the agent is practically

irrational. Practically irrationality rather should be understood as a *conative* failing. That this is what Smith intends is clear by his appeal the much cited article by Michael Stocker, "Desiring the Bad: An Essay in Moral Psychology," to elaborate upon the circumstances in which the connection between moral judgment and motivation can be expected to break down. Smith quotes Stocker at length:

> Through spiritual or physical tiredness, through accidie, through weakness of body, through illness, through general apathy, through despair, through inability to concentrate, through a feeling of uselessness or futility, and so on, one may feel less and less motivated to seek what is good. One's lessened desire need not signal, much less be the product of, the fact that, or one's belief that, there is less good to be obtained or produced, as in the case of a universal Weltschmerz. Indeed, a frequent added defect of being in such 'depressions' is that one sees all the good to be won or saved and one lacks the will, interest, desire or strength.[26]

Smith's immediate commentary on this passage leaves no doubt about the appropriateness of a conative construal of 'practically irrational':

> It is a commonplace, a fact of ordinary moral experience, that *practical irrationalities* of various kinds—various sorts of 'depression' as Stocker calls them—can leave someone's evaluative outlook intact while removing their motivations altogether (MP: 120-1, emphasis added).

According to the proper interpretation of PR, the immorality of the inappropriately motivated agent is due to the agent's practical irrationality, where 'practical irrationality' implicates faulty *conative*, rather than cognitive, capacities.

Nevertheless, it is not obvious that a conative interpretation of practical irrationality is sufficient to render PR consistent with MR. We need a deeper grasp of the conative dimension of practical rationality. Recall that the statement of MR given above included, essentially, the qualification that the immoral agent's inappropriate motivation could not be due to some psychological disorder or dysfunction for which the agent was not responsible

---

[26] This is quoted on page 120 of *The Moral Problem*. Stocker's article was originally published in *The Journal of Philosophy*, Vol. LXXVI, No. 12 (1979): 738-53. Quotation is from page 744.

nor beyond her ability to overcome. The point of such a qualification is clear: that an agent's inappropriate motivation is due to such a disorder or dysfunction serves, according to our moral practice, to undermine the idea that the agent is responsible. It becomes doubtful that the reactive attitudes involved in holding an agent responsible are appropriate if that agent's motivations are beyond her control or, in some sense, 'not her own.' In light of this we need to give rather careful consideration to the explication of irrationality Smith gives in his discussion of the motivational force of normative judgments. Smith claims that

> Desires are irrational to the extent that they are *wholly and solely* the product of psychological compulsions, physical addictions, emotional disturbances and the like: to the extent that they wouldn't be had by someone in a non-depressed, non-addictive, non-emotionally disturbed state (MP: 155, emphasis in the original.)

There is something obviously intuitive about this unpacking of the concept of (practical) irrationality. It captures an important feature of our commonsense deployment of that concept: motivation is irrational when it is not in the control of the *agent* but rather the result of some distortion of their capacities, a distortion that, presumably, the agent is not responsible for. But though this account of irrationality is perfectly consistent with our intuitions on the matter it is not at all obvious that it can function in an *explanation* of *immorality*. If an agent's motivations are "wholly and solely" the product of such conditions and, further, the agent is not responsible for those conditions obtaining, it is quite unlikely that that agent will be held morally responsible for those motivations. Indeed, many would object to holding such an agent responsible on *moral* grounds: the morally inappropriate attitudes and motivations in such a case would belong to the evaluators, not the evaluated. So understood, practical irrationality does not so much explain immorality as excuse it.[27]

---

[27] The example of Adrea Yates, the Texas woman who drowned her five children in the bathtub, comes to mind. Much of the debate, among the public at least, concerned whether her actions were due to some psychological condition that put her behavior beyond her control. If so, most thought she should not be found morally (or criminally) responsible (which is not to say that she should not be removed from the public).

Actually, the situation is worse than this. In giving an explication of practical irrationality in terms of psychological disorder, the weak internalist is in danger of rendering the notion of moral culpability thoroughly problematic, if not incoherent. If the only circumstances wherein an agent fails to be motivated appropriately with respect to her moral judgments are precisely those in which she suffers from some compulsion, disorder, or disturbance then it becomes questionable how MR can be true.

The status of MR as a conceptual truth demands that there be failures of moral motivation that are *not* wholly and solely the product of psychological compulsions, disturbances, and the like. And perhaps there are such instances of motivational failure. Perhaps the weak internalist can defend her position by appealing to the very qualification in MR. If the irrationality that accounts for the inappropriate motivation is something for which the agent herself is responsible and for which she can be blamed, then PR, appropriately qualified, could be made compatible with MR. The idea would roughly be this: if an agent knowingly and willingly acts in ways that render her physically addicted, or emotionally disturbed and, as a result, she fails to be appropriately motivated with respect to her moral judgments, then she would be immoral *because* she was practically irrational. And aren't we in fact confronted with many instances of this? Alcoholics, drug addicts, and abusers of various kinds who have, of their own free choice, rendered themselves morally bankrupt would seem to be tailor-made examples for the weak internalist. The core idea is an old one: though, having formed a bad (irrational) character leaves us unable to choose otherwise than we do, we are responsible for the choices that led to the formation of that character.[28] Perhaps, then, PR and MR *are* consistent after all.

---

[28] See Aristotle, *NE* III 5 1114a 3-22.

Old and venerable as this type of thinking may be, it is not obvious that the weak internalist can avail herself of it. In order to see this we need to make clear precisely what the weak internalist position involves. What is being suggested is that, in order to make both PR and MR true, we see the failures of motivation covered by PR fall into two classes: those for which the agent is legitimately held responsible for her irrationality and those for which she is not. Such a distinction would have no bearing upon *rational* evaluation, since such evaluation does not discriminate between failures of rationality for which the agent is responsible from those for which she is not: the agent is deemed irrational in either case. But such a distinction is essential for the purposes of moral evaluation, for moral evaluation *must* discriminate between moral failures for which the agent is responsible from those for which she is not. Indeed, failure to do or be motivated as one morally ought that is not due to something for which the agent is responsible is arguably not an instance of *moral* failure at all. Any attempt, like that of the weak internalist, to make immorality a species of irrationality, must make this distinction. If immorality is to be understood in terms of irrationality then it must be something akin to 'agent-induced' irrationality. The problem is that this idea does not make much sense on weak internalist assumptions.

The weak internalist's explanation of inappropriate motivation in terms of irrationality is grounded in the idea that moral judgments are a species of rational judgment: what is morally appropriate is what is rationally appropriate; what we morally ought to do is what we have good reason to do (MP: 91,96). Again, the idea is clear enough: assume that judgments of rationality are at once objective and practical, assume that moral judgments are a species of rational judgments, and then the core thesis of weak internalism falls out: moral judgments are at once objective and practical (MP: Ch. 5; D: 88-107). Rational judgments "express norms of rationality or reason," telling us "what is and is not rational for us to do"

(MP: 130). Such judgments provide 'normative reasons for action,' where a 'normative reason' is indicative of the relevant action's *desirability*. As Smith puts it:

> [I]t is a platitude to say that what it is desirable that we do is what we would desire to do if we were fully rational; that what we have a normative reason to do is what we would desire that we do if we were fully rational (MP: 150)

Platitude or not, this conception of rational judgments provides the basis for a conception of moral judgments, a conception that Smith describes as "expressions of our belief about what we have normative reason to do, where such reasons are in turn categorical requirements of rationality" (MP: 185). Now for our present purposes, the crucial claim is that such moral *beliefs* (such judgments are meant to be objective) are practical in virtue of the practicality of rational judgments. The practicality (i.e. motivational influence) of rational judgments amounts to the following: when an agent believes that she would desire to perform some action if she were fully rational she then acquires the desire to perform that action (if she were not already so disposed) on pain of being irrational. Since moral judgments are a species of rational judgments, PR drops out. But PR needs to be compatible with MR, and the suggestion we are presently considering is that failures of motivation that qualify as immoral—those which are blameworthy—are instances of motivational irrationality for which the agent is responsible, what we are calling 'agent-induced' irrationality. The problem, however, is that 'agent-induced' irrationality is beyond the weak internalist's scope: the very 'practicality' of rational judgments undermines the idea that an agent could be responsible for any irrationality.

The weak internalist claims that an agent is motivated to do what she judges she is rationally required to do or she is irrational. This is a conceptual truth. If an agent judges that she would have some motivation if she were fully rational and she does *not* have such a motivation then she is irrational and irrational "by her own lights." We have already been

told what would account for such dissonance between rational judgment and motivation: the agent's actual motivations being "wholly and solely the product of psychological compulsions, physical addictions, emotional disturbances and the like" (MP: 155). Now if 'agent-induced' irrationality is a plausible notion, one sufficient to ground judgments of immorality, it must be the case that an agent can bring about such compulsions, addictions, and disturbances *and in a manner for which she can be held morally responsible.* But it is rather hard to fathom how an agent *could* be responsible for inducing the type of psychological dysfunction and disorder that is indicative of an irrational condition if agents are necessarily motivated to do what they think is rationally required (absent some psychological disorder). For who would possibly think that it was rationally required to do something that would render her psychologically disordered? Presumably only those who either already suffer from some such disorder, or who are ignorant of the consequences of their proposed actions, or who have radically false beliefs about what is rationally required. However, our moral practice strongly inhibits holding any of these types of person responsible for their choices and behavior in this regard. Such people deserve our pity and guidance, not our blame: such people do not *intend* to render themselves irrational, they are not motivated by that goal (at least not under that description). For an agent to be responsible for her irrationality she must knowingly and willfully cause, or do something that causes, or results in her irrationality. But this does not seem possible if rational agents are necessarily motivated to do what they believe to be most rational.

'Agent-induced' irrationality is not a tenable account of immoral motivation. In looking for a diagnosis of this account's failure it seems that Hume's point is still pertinent: for an agent's motivation to be (morally) evaluable basic conditions must be satisfied, conditions that render the agent (and her motivations) appropriate objects of moral address.

We are seeing, as Hume suggested, that what is perhaps the most essential condition that the agent must satisfy is that she qualify as rational. We have already seen that being rational must involve the cognitive ability to *make* judgments, especially judgments of value. But we are now seeing that being rational must also involve the ability to make choices, act (or not act) in virtue of the judgments one has made and, ultimately, *be* motivationally oriented in unrestricted ways in (relatively) *normal psychological* conditions. That is, being (practically) rational involves what we might call 'conative autonomy.'[29] The reason why motivations that are "wholly and solely" the product of dysfunction and disorder are irrational is precisely because they are *not* autonomous—the disorder was not authored by the agent herself. But the very coherence of moral culpability requires that the inappropriateness or the lack of moral motivation *be* authored by the agent. That is what immorality consists in. When, however, problematic motivation is not the agent's doing, we employ a different term. Such agent's motivation is not immoral but irrational. Immorality cannot be explained in terms of practical irrationality, because practical rationality—conative autonomy—is one of its preconditions.[30]

---

[29] It should be noted that the notion of 'autonomy' being used here, that of being capable of being motivationally 'unrestricted' in normal psychological conditions is not one that is metaphysically loaded in the sense of requiring some 'libertarian' or Kantian sense of freedom; what might be called 'contra-causal' freedom. It is only meant to indicate that for an agent to be held responsible her responsiveness to reasons for actions (which include the type of action, its possible consequences, etc.) cannot be necessitated by the very content of those reasons. It cannot be, for instance, that an agent who normally responds to certain reasons for action can fail to respond to such reasons *only if* she were to be suffering from some psychological dysfunction of some kind. It should not be too much to require that agents who are cognizant of reasons that are normally quite significant to them be able to say 'not today' in order to see them as morally responsible agents. I take this to be very near in spirit to the position Christine Korsgaard argues for in her "The Normativity of Instrumental Reason," *op. at.*, though I am not at all confident that I am using it argumentatively in the manner she would find most appropriate. For an very different account of the relationship between responsibility, acting rationally, and acting autonomously, see Susan Wolf, "Asymmetrical Freedom," *The Journal of Philosophy*, Vol. LXXVII, No. 3 (1980): 151-66; and "The Reason View," from her *Freedom Within Reason* (Oxford: Oxford University Press, 1990), reprinted in Laura Waddell Ekstrom's edited *Agency and Responsibility: Essays on the Metaphysics of Freedom* (Boulder, CO: Westview Press, 200): 205-26.

[30] Again, it is worth keeping an example like Andrea Yates in mind here. It is generally agreed that she was, at the time immediately leading up to the murder of her children, in a psychologically disturbed state of mind. The question that people have struggled with is whether she was psychologically capable of refraining from

The weak internalist's construal of *rational* motivation (and, in turn, moral motivation), if correct, would put cognitive autonomy itself in doubt. For agents who are necessarily motivated in accordance with their rational judgments are *not* conatively autonomous.[31] According to the weak internalist, an agent's motivations are determined, not by the agent herself, but rather by the content of the agent's judgments. The passivity this model of motivation involves must cause us to doubt whether the creatures it ranges over are in fact rational *agents* at all. For on this model, if an 'agent' does *not* suffer from some psychological dysfunction then she *must* be motivated in accordance with her rational judgments. But this is unacceptable. *Bona fide* agency must allow for a psychologically normal agent *choosing not* to pursue what she judges would be the most rational (or, for that matter, the most moral) thing to do. Anything less threatens the very coherence of the moral evaluation of agents, their motives, and actions. The weak internalist's explanation of immorality in terms of irrationality is untenable. Immoral agents *must be* rational.

Immorality as Irrationality, Part II

It would seem a methodological dictum that we not employ explanations of phenomena that render those very phenomena impossible or incoherent.[32] The phenomena themselves must dictate where we look for adequate explanations. And this should make us wonder about

---

acting on the thoughts and impulses she had. If she was capable yet acted anyway we feel that we have no recourse but to condemn her as immoral. If, on the other hand, she was not capable of resisting, if her condition removed resistance as an option, we feel it natural to describe her as irrational, though not necessarily as immoral.

[31] Note that this point applies to *any* account that both defines being rational in terms of being necessarily motivated in accordance with judgments of rationality *and* claims that moral judgments are to be construed as a species of rational judgments, as weak internalists do.

[32] Of course we are sometimes in the business of *explaining away* certain phenomena, but this is not what the weak internalist is up to. To think so is to confuse her with the moral skeptic. Weak internalists are, after all, moral *realists*: immorality is a real phenomenon.

the wisdom of employing the very framework of the explanation we have been considering, that of understanding morality in terms of rationality. The moral realm, in a very important sense, *presupposes* rationality, but it should not be conflated with it. Doing so leads to the assimilation of immorality to illness and this has the air of a category mistake. Moral significance—positive and negative, good and bad—obtains *within* the bounds of rational significance. But it is important not to misunderstand this point. Morality presupposes rationality, but *not* in the sense that the former is a subset of the latter. Rationality, in the sense of being able to reflect rationally *and* in the sense of exercising conative autonomy, provides morality with its context, not its content. Moral judgments are not judgments about what is most 'rational' to do. In this section I want to further probe the failure of the rationalist framework the weak internalist employs by looking at the idea that moral judgments are based on, or are a species of, rational judgments in greater detail. Building upon the conclusions of the discussion so far, I aim to illuminate the difficulties that the weak internalist faces and to explore further the nature of the relationship between morality and rationality.

The false notes of the weak internalist's rationalist paradigm can be heard plainly when we focus on its core concept: the 'fully rational agent.'[33] The weak internalist, as we have seen, explains the putative psychological link between moral judgment and motivation by explaining it in terms of being practically rational: moral judgments are a species of rational judgments and a practically rational agent is, by definition, motivated in accordance with

---

[33] This idea receives its most extensive treatment in the work of Smith, particularly his "Dispositional Theories of Value," *Proceedings of the Aristotelian Society*, Supplementary Volume (1989): 89-111; *The Moral Problem*, *op.cit.*; and "Internal Reasons," *Philosophy and Phenomenological Research*, Vol. LV, No. 1, (1995): 109-31. The idea is also developed in Richard B. Brandt's *A Theory of the Good and the Right* (Oxford: Oxford University Press, 1979; reprinted by Prometheus Books, Amherst, New York, 1998); Bernard Williams' "Internal and External Reasons," reprinted in his *Moral Luck* (Cambridge: Cambridge University Press, 1981); and by Christine Korsgaard, "Skepticism about Practical Reason," *The Journal of Philosophy*, *op. Cit.*, 5-26. I will focus here on

such judgments. The significance of the *fully* rational agent is to posit an ideal that less than fully, or imperfectly rational agents such as ourselves are advised to consult for answers to our most vexing practical questions. The fully rational agent is a model that we are urged to pattern ourselves after to achieve practical success. But since moral judgments are actually rational judgments, the fully rational agent is the gold standard, not simply of rationality, but of morality as well: to be fully rational is to be thoroughly virtuous. 'Full rationality' embodies not only perfect clarity in thought and reasoning but an unshakeable moral certainty, resulting from an unerring and infallible sensitivity. But this account of full rationality seems deeply problematic and without much in the way of justification. 'Full rationality' does not seem to be a notion sufficiently stable and coherent to support the weak internalist's conception of practical psychology and its moral relevance.

It is crucial to note immediately that the idea of a fully rational agent as one who is fully virtuous, or unerringly motivated in accord with her moral judgments, is purely stipulative.[34] How, precisely, 'fully rational' is best understood and theoretically deployed is quite a contentious point. Those who do not share the view that morality is a species of rationality would undoubtedly question the combination of attributes of impeccable reasoning and moral virtue. Why the fully rational should be thought of as necessarily virtuous is not at all obvious to one who has not already accepted the identification of morality with rationality. To the unconverted being completely rational is one thing, being good another. The appropriateness of a stipulative definition is, of course, a matter of how well it accounts for our most deep-seated intuitions—its explanatory utility and adequacy.

Smith's development of this notion, though I hope that it is clear that the criticisms offered here are meant to apply to the position in general, and not simply to a particular version.

[34] As Smith himself concedes. In responding to an objection to his use of 'rationality,' Smith defends his usage by stating that "the term 'rationality' is almost entirely a philosopher's term of art" (D: 91).

Intertwining perfect intelligence with perfect virtue is, of course, not a new idea. It is obviously present in the extreme rationalism of philosophers like Socrates and Plato, and, to a somewhat lesser extent, Aristotle. The amalgamation has also proved attractive to the more theistically minded. Indeed, the fully rational agent possesses characteristics strikingly similar to the God of the New Testament. That version of the Supreme Being is not only omniscient (and infallible in His reasoning)[35], and omnipotent, but also omnibenevolent—He is all-good. Now fully rational agents, are not thought of as all-powerful, but they are endowed with other perfections—on a human scale, of course—that make them exemplars. For instance, in Smith's terms, the fully rational agent can be identified by the following criteria:

A. No physical impairments (that could disturb reasoning)
B. No emotional impairments
C. No false beliefs
D. Possesses all relevant true beliefs
E. Possesses a systematically justifiable motivational set (that is, a 'maximally coherent and unified' motivational set)[36]

Fully rational agents, though not omniscient, believe enough true propositions to be completely apprised of their situation, their options, and capable of impeccably calculating all of their ramifications. This would seem part and parcel of *everyone's* idea of the 'fully rational.' But whither their goodness? Why think the aforementioned psychological virtues will bring moral virtue in their train? It is worth noting that the goodness of God is best understood not as a product of His *knowledge* (to assume so is to forfeit a Divine *grounding* of value) but of His *perfection*: a perfect being could be nothing less than perfectly

---

[35] Though whether God *does* in fact reason is questionable. That is to say that God's knowledge can be thought to be in no way dependent upon inference—he simply knows. Of course, we would expect that he *could* reason with the best of them, if He were so inclined.

[36] This account of the fully rational agent is taken from page 89 of "In Defense of *The Moral Problem*: A reply to Brink, Copp, and Sayre-McCord."

good and His creations necessarily possess an inherently perfect order (the natural/moral order). But this rationale is, of course, debatable (the goodness of God might strike us as a matter of definition, after all, not of discovery). Needless to say, it does not obviously betray a lack of mastery of the concept of God to wonder whether what He does and would do is always morally right; whether His motives are always virtuous. Should those kinds of question then call into question our mastery of the concept of the fully rational agent? In any event, what reason do we have for thinking that the concept of the fully rational agent will prove any more theoretically useful than its theistic analogue?

The idea of grounding moral judgment in terms of the judgments of the fully rational agent should strike us as a non-starter. The very point that suggests that it would be useful threatens to undermine it completely, namely the view that we are, at best, only imperfectly rational and that it is such imperfection that is responsible for the moral failings we have.[37] Our imperfect rationality stands in stark contrast to the criteria embodied by the fully rational agent outlined above. Whereas fully rational agents possess all relevant true beliefs, no false beliefs, and a maximally coherent and unified motivational set, actual agents like ourselves labor under all kinds of burdens and limitations. What is perhaps maddeningly indicative of the human condition is that an individual's motivations, taken in their totality, are internally inconsistent. This is, in large part, the reason why we have practical difficulties in the first place.[38] It is precisely this inconsistency that provides the basis for practical reflection. But such reflection is necessarily circumscribed: given our limitations with

---

[37] This, I am suggesting, is the view that gives impetus to the idea that full rationality involves full virtue. There is, however, no reason to think that full rationality is either a necessary or sufficient condition for full virtue. Though being imperfectly rational does, as a matter of fact, provide us with the circumstances of our immorality, it does not preclude our moral development. Imperfectly rational beings could be perfectly virtuous.

[38] Of course environmental contingencies play their crucial role as well.

respect to information, computational power, processing speed, and life-span, our rational concern with the consistency of our motivations (and, in turn, choices and actions) must be parochial. Global consistency is not something we are likely to have even the vaguest ideas about; it is thoroughly beyond us. Our focus must be a local one if we are to have anything like a sufficiently clear picture of our situation and what we might possibly do about it.

Given these limitations, the idea of the fully rational agent seems to be of little use for the purposes of normative reasoning. The problem is not that there are no such agents—the advocate of the fully rational agent is not claiming that there are, but only offering it as an ideal—but that imperfectly rational agents, such as ourselves, haven't the slightest idea what a fully rational agent would believe or be motivated to do. We can't imagine what it is like to have only true beliefs except to say that we would have only true beliefs—we have no idea what their content would be. Nor does the idea of a 'maximally coherent and unified' motivational set have much in the way of content beyond the impressionistic picture of a system of motivations that has succeeded in minimizing motivational conflict (most likely *via* the imposition of a preference ordering). We simply lack the information—and the requisite ability to process it—to flesh out the idea sufficiently to render it theoretically interesting. Asking ourselves what *we* would be motivated to do if *we* were fully rational is akin to asking a personal computer circa 1985 to simulate, in real time, a present day Cray super-computer: it can't be done. Since 'should' implies 'can', the fact that we can't implies that we needn't.[39]

Trying to divine the motivations of the fully rational agent is, for creatures such as ourselves, akin to understanding the mind of God, which is inscrutable: we can't know them.

---

[39] I wish to thank Gerrit Jan Kamperdyk for impressing upon me the importance of the argument of this paragraph, as well as for the example.

If moral judgments are judgments of what we would be motivated to do if we were fully rational, then we cannot possibly know what is right or wrong, appropriate or inappropriate. In light of this difficulty, arguments concerning whether *our* idea of what a fully rational agent is motivated to do is relative to the actual motivations that *we* have or if we would all *converge*, and necessarily so, on the motivations that all fully rational agents supposedly share, are completely beside the point.[40] Since we can have no idea what motivations a fully rational agent would have, the question of whether two fully rational agents are necessarily motivated in the same ways or not is unanswerable.

Smith, for his part, claims that "we do not need to make any extravagant assumptions about our psychological powers" in order to determine what motivations we would be justified in having as fully rational agents (D: 106). Instead, Smith claims

> We need simply to be able to engage in the imaginative process of making our desires maximally informed, unified, and coherent, and to be disposed, in so doing, to respond appropriately should we discover that the desires that others come up with when they too engage in this imaginative exercise are different from ours. We need to see ourselves as in disagreement with these others and to be willing and able to provide further arguments in support of our own desires, or else to change our desires in response to the arguments that they offer in defense of theirs. Far from these being extraordinary psychological abilities, they are very ordinary abilities, abilities that we take each other to have as a matter of course in the give and take of everyday life (D: 106-7).

Whether such psychological abilities and efforts should be thought extraordinary or not, there is something troubling about Smith's account of the 'appropriate' response to disagreement about the results of this imaginative process. Granted we are expected to think more deeply and offer more arguments for what we think is the correct answer about what our motivational profile would look like if we were fully rational (and, of course, concede the point when we think we have been bested by someone who's imaginative

---

[40] See, for instance, the criticism of Bernard Williams' 'Humean' conception of the fully rational agent by Smith in *The Moral Problem*, 164-74, and in "Internal Reasons," 117-25.

powers possess greater accuracy then our own). But what happens when two opposing contemplators are unwilling to back down, both thinking that their arguments are superior to those of their interlocutors? What is the 'appropriate' move at this point?

It seems to me that what we expect at this point is precisely what we in fact do as a matter of course in the give and take of everyday life, namely we agree to disagree. What the most rational thing to be motivated to do, according to one person, simply need not be what some other person thinks would be the most rational thing to be motivated to do. And this seems perfectly adequate so long as these different people with their different ideas about what is most rational can peacefully co-exist. If, however, the matter of disagreement is such that there can be only *one* policy, or course, for a group of deliberators to adopt there seems little reason to think that chosen course will be determined by a process of rational argument. Indeed, in such situations things are often decided by vote, or by the acquiescence of one or the other side for the purposes of peaceful relations, or by the brute force of the more powerful. Rarely do those who initially held opposing views about what would be the most rational course abdicate their views as a result of being persuaded by reasoned criticism. There seems woefully little precedent for thinking that there is a 'most rational' course of action or motivation that all 'reasonable' people will recognize. What is rational to be motivated to do seems to be irredeemably relative to the motivations that agents in fact have when they engage in the process of reasoned criticism. But if this is correct, the attempt to ground moral judgments in rational judgments is antithetical to moral realism. However the realist position might be cashed out, it cannot be relativistic. The weak internalist approach appears deeply flawed: if moral judgments are properly understood in terms of judgments about what is most rational then those judgments would not seem to

be objective. If moral judgments are objective then their putative practicality cannot be secured by assimilating those judgments to judgments of rationality.

There is, I believe, a further, more troubling aspect of the assimilation of moral judgments to judgments of rationality that deserves our attention. We will recall that the proposed assimilation was meant to serve an explanatory function: moral agents are such that they are motivated to do what they judge they ought to do or they are immoral *because* rational agents are motivated to do what they judge would be most rational to do (otherwise they would not be rational but rather irrational) *and* moral judgments are a species of rational judgments. As noted, this rationalist explanation of morality and moral phenomena finds expression in PR:

> PR: If an agent judges that it is right to ϕ in circumstances C, then either she is motivated to ϕ in C or she is practically irrational.

We saw in the last section that for this rationalist explanation to be successful two conditions must be satisfied. The first is that the truth of PR must be consistent with the truth of MR. It cannot be the case that the *moral* fact that agents who are inappropriately motivated with respect to their moral judgments are immoral is *undermined* by the claim that those who are so motivated are practically irrational. The second condition on the adequacy of the rationalist explanation is that it be able to defend the claim that moral phenomena, such as that embodied in MR, is in fact *dependent* on rationalist phenomena, such as that embodied in PR. This second condition is merely the demand that the framework of practical rationality be *explanatory* of moral phenomena. But it is one thing to *claim* that rationality can explain morality and quite another to actually do so.

We argued at length in the previous section that the first condition of adequacy for PR cannot be satisfied: there does not seem anyway of construing PR such that it is consistent

with the truth of MR. In the remainder of this section I want to argue that we also have no reason to think that the second condition can be met. Indeed, the difficulty here may be such that the satisfaction of the first condition of adequacy would illustrate the hopelessness of using PR to explain MR. The difficulty is this: *if*, contrary to what we have argued, agents could be held accountable for their irrationality, *then* their very irrationality is an appropriate object of evaluation. But now the defender of weak internalism faces the challenge of explaining why irrationality has the moral value that it does. Why, that is, is such irrationality immoral? That we in fact find such (intentional) irrationality immoral might be granted, but this only serves to suggest that there is an antecedent evaluative (namely moral) framework according to which such irrationality in some way falls short. And if there is such an antecedent evaluative framework then morality would not be dependent upon the rationalist framework and therefore the latter could not actually explain the former. Indeed, it might be case that some instances of willful irrationality (in the sense of failing to be motivated in the way one judges one would be if one were rational) are due to the *immorality* of the agent. By allowing agents the conative autonomy sufficient to render them responsible for their irrationality—and thereby making a principle like PR consistent with MR—the weak internalist may well rob irrationality of the power to explain immorality.

We can see the bind the weak internalist is in by considering the following passage from Smith.

> Consider those who believe that they would desire that they keep a promise in circumstances C if they had a maximally informed and coherent and unified set of desires, and who also desire that they keep a promise in C, and compare them with another group of people who have the belief but lack the desire. It seems to me that the first group plainly has a psychology that, in this respect at any rate, exhibits more in the way of coherence than the latter. There is a disequilibrium or dissonance or failure of fit involved in the latter psychology, where there is equilibrium or consonance or fittingness involved in the former. Rationality, in the sense of this sort of coherence, is thus on the side of agents whose desires match their beliefs about the normative reasons that they have. Exhibiting this

> sort of coherence is what practical rationality consists in, and failing to exhibit it is what practical irrationality consists in.
>
> If this is right, however, note that we not only have an explanation of why people who fail to desire what they believe they have a normative reason to do are practically irrational, but that we also have an explanation of the mechanism by which practically rational agents can come to desire to do what they believe they have a normative reason to do. Agents whose psychologies exhibit this sort of coherence—that is, those whose psychologies exhibit a tendency toward such coherence—will have desires that match their beliefs about their normative reasons, whereas those whose psychologies do not exhibit this sort of coherence may fail to have such desires. Beliefs about normative reasons, *when combined with an agent's tendency to have a coherent psychology*, can thus cause agents to have matching desires (D: 100, emphasis added).

The dilemma facing this analysis is quite clear. On the one hand, if the tendency towards coherence is *not* in the control of the agent then it is not something for which we can legitimately hold the agent responsible. The problem is that *moral* criticism of a dissonant psychology *requires* that we are able to hold the agent accountable for having or failing to have such a tendency. Failure to have such a tendency where that failure is due to some psychological dysfunction *would* serve to explain why an agent was inappropriately motivated in accordance with her moral judgments but it would preclude that failure from counting as immoral. On the other hand, if we *can* hold an agent responsible for the coherence, or lack thereof, of her psychology—if the psychological mechanisms behind this sense of rationality and irrationality are things for which we *can* hold an agent accountable—we are then owed an account of why the lack of such coherence (or the lack of a tendency toward such coherence) should count as being immoral. If the motivations that derive from such an incoherent psychological profile are to count as immoral then we should expect that the *immorality* of those motivations should likewise derive from such incoherence. The problem is that we have no reason to believe that it does.

Recall our earlier discussion of the idea of 'agent-induced' irrationality. The difficulty we highlighted there was that, given the weak internalist's insistence that rational agents are

necessarily motivated in accordance with their normative judgments, rational agents are not sufficiently autonomous to induce irrationality. We are now putting such difficulties aside, however, and allowing rational agents the requisite autonomy to be the authors of the dissonance between their motivations and their normative judgments. We are allowing that is, that rational agents can be held responsible for their 'practical irrationality', where that is understood in terms of psychological dissonance. But to concede this is still not enough to *explain*, or *ground* the dissonance of motivation and moral judgment that is constitutive of immorality in terms of practical irrationality. Further argument would be required for that. But the argument that Smith provides, that it is a 'platitude' about normative judgment (and, in turn, moral judgment) that it concerns what we would be motivated to do if we were fully rational, won't work. The idea of the 'fully rational agent' is without content: we really have no idea what a fully rational agent would be motivated to do.

Admittedly the argument being presented here makes liberal use of a platitude about morality, namely that moral motivation (and, of course, conduct) is something for which agents are responsible and which therefore can be praised or blamed. But this platitude is, I submit, far more secure than the claim that moral judgments are judgments of what we would be motivated to do if we were fully rational. It is also more secure than putative platitudes such as 'Whether or not ϕ-ing is right can be discovered by engaging in rational argument' or, 'Provided A and B are open-minded and thinking clearly, an argument between A and B about the rightness or wrongness of ϕ-ing should result in A and B coming to some agreement on the matter,' platitudes that Smith claims the correct analysis of morality will capture.[41] But these do not seem to be *platitudes* of moral thought at all, but

[41] *The Moral Problem*, 39-40.

rather tendentious assumptions about the proper form moral theory ought to take. For if these were in fact platitudes then we ought to be able to reject unequivocally those who deny that moral disputes can be resolved by rational argument by simply pointing out their obviously incomplete grasp of the moral terrain. But such 'platitudes' cannot be so employed because, in large part, it is precisely such claims that are in dispute among competing metaethical positions. There are precious few theory-neutral platitudes that can be assumed in theory construction. But without the support of such platitudes there seems no reason whatever to think that the intentional inducement of irrationality on the part of the agent captures anything essential about immorality, such that irrationality could serve to explain immorality.[42]

The worry here for the weak internalist is a real one and it is nicely expressed in Hume's critique of rationalism. Consider the following related criticisms of the explanatory capacity of the rationalist paradigm.

> Should it be pretended that tho' a mistake of *fact* be not criminal, yet a mistake of *right* often is; and that this may be the source of immorality: I would answer, that 'tis impossible such a mistake can ever be the original source of immorality, since it supposes a real right and wrong; that is, a real distinction in morals, independent of these judgments. A mistake, therefore, of right may become a species of immorality; but 'tis only a secondary one, and is founded on some other, antecedent to it.[43]

> But what may suffice entirely to destroy this whimsical system is, that it leaves us under the same difficulty to give a reason why truth is virtuous and falsehood vicious, as to account for the merit or turpitude of any other action. I shall allow, if you please, that all immorality is derived from this supposed falsehood in action, provided you can give me any plausible reason, why such a falsehood is immoral. If you consider rightly the matter, you will find yourself in the same difficulty as at the beginning.[44]

---

[42] For further criticism of the 'platitude defense' of weak internalism, see Shuan Nichols, "How Psychopaths Threaten Moral Rationalism: Is it Irrational to be Immoral?" *The Monist* Vol. 85, No. 2 (2002): 285-303.

[43] *Treatise of Human Nature*, 460, emphasis in the original.

[44] Ibid., 462.

Though Hume was ostensibly arguing against a different 'rationalist' conception than we are presently considering, the point is perfectly general. One needs only substitute 'coherence' and 'lack of coherence' for 'truth' and 'falsehood' and the relevance of the argument becomes plain.[45] What plausible reason do we have for thinking that failure to be motivated in accordance with our judgments of what is most rational to do is *immoral*? Without the support of honest-to-goodness platitudes about the rational basis of morality we appear to have no plausible reason whatsoever for thinking that irrationality is necessarily immoral. Indeed, we can come up with examples that show that our thoughts about irrationality and immorality can *diverge*.

Take, for instance, the case of Huckleberry Finn. As many philosophers have pointed out, Huck clearly possesses motivations that are inconsistent with his judgment about what is morally right.[46] Though he thinks that he ought to surrender his friend Jim to the slave hunters his affection for his friend gets the better of him and he refuses to turn Jim in. Further, let it be the case that he not only thinks that it would be right to turn Jim in but that it would be most rational to do so. Huck is then, if we wish, practically irrational.[47] But is he immoral? I think most of us would say that he is not, but is rather praiseworthy for following his better nature. Does this moral judgment of ours concerning Huck reveal that *we* think that slavery is wrong and that his judgment that it is morally right to turn Jim in is

---

[45] Again, such criticism should be seen to be applicable to *any* account of rationality that is meant to serve as the basis of morality, not simply Smith's.

[46] For instance, Jonathan Bennett, "The Conscience of Huckleberry Finn," *Philosophy*, XLIX (1974): 123-34; Alison MacIntyre, "Is Akratic Action Always Irrational?" in *Identity, Character, and Morality: Essays in Moral Psychology*, edited by Owen Flanagan and Amelie Oksenberg Rorty (Cambridge: MIT Press, 1990): 379-400; Thomas Hill, "Four Conceptions of Conscience," *Nomos*, LX (1998): 13-52; and Nomy Arpaly, *Moral Worth*, *The Journal of Philosophy*, XCIX, No. 5 (2002): 223-245.

[47] Though see McIntyre (ibid.) for arguments to the effect that Huck is *not* practically irrational.

mistaken? I have no doubt that it does. But notice that we can take this stand about Huck's judgment about what is right while happily conceding that Huck's judgments about what is most rational to do are perfectly fine. Turning Jim in was, arguably, the most rational thing for a child like Huck (or anyone, for that matter) to do. But its being the rational thing to do does not make it the morally right thing to do. Nor does the psychological incoherence that Huck thereby exhibits in not being motivated to do what he thinks he would be motivated to do if he were fully rational show that he is immoral. Being irrational and being immoral are not the same.

Our commonsense judgments concerning rationality and morality reveal that we take these to be independent evaluative frameworks. It is not, therefore, a platitude that we take moral judgments to be grounded in judgments of rationality. Nor does our commonsense views give us any reason to think that irrationality can function as an adequate explanation of immorality. On the contrary, appeal to our commonsense views suggests that we have in morality an evaluative framework that, however recondite, is *already* in place and which can then be brought to bear on failures of rational motivation. In some cases, like that of Huck Finn, such failure will be praised, whereas in many others the failure to be motivated appropriately with respect to what is judged most rational will be condemned. Others still, will escape moral judgment altogether, being outside of the sphere of proper moral address. These last are the truly irrational, those who's psychological incoherence is beyond their control.

Moral Indifference, moral goodness

Our argument that weak internalism is a confused moral psychology and that cognitivists about moral judgment ought to be externalists about motivation would not be complete if

we did not address what Smith takes to be the primary argument against externalism. Externalists have criticized internalism on its inability to either allow for or adequately explain the *amoralist*, the person who competently makes moral judgments but who is unaffected by them; in other words, the morally indifferent agent. Since internalism claims that moral judgments are necessarily motivational, either intrinsically or relationally, the idea of an agent who is unmoved by her moral judgments makes little sense. The most that the internalist could say about such a person is that though the amoralist may exist, she is necessarily practically irrational.[48] But this answer, given what we have said above, won't do. An agent may be indifferent with respect to her moral judgments, and be so *without* being practically irrational. Indeed, if the amoralist's indifference *is* to be explained by her being irrational, then she is no longer an instance of a *moral* category: she would not be morally responsible for her indifference.

This is something for externalists to keep in mind, no less than internalists. Some discussions of amoralism tend to speak somewhat interchangeably between amoralists, on the one hand, and *psychopaths* and *sociopaths* on the other.[49] Psychopaths and sociopaths are controversial figures in moral discussions, not because it is questionable whether they exist, but rather that it is questionable whether it is appropriate to hold them morally responsible for their behavior. If the answer to that question is that it is not, then such persons are not, properly speaking, *immoral*. Our previous discussion of irrationality, however, gives us a diagnosis of the difficulty such people present. The question whether psychopaths should be punished or pitied has much to do with the question of whether they exhibit conative

---

[48] See Smith, "The Argument for Internalism: reply to Miller," *op.cit.*,176-8.

[49] See, for instance, Smith, *The Moral Problem*, 67, and Shuan Nichols, "How Psychopaths Threaten Moral Rationalism: Is it Irrational to be Amoral?" *op. at.*

autonomy or not. Those who do are candidates for retribution; those who do not are candidates for benevolence.

The truth is that we need the culpably indifferent, just as much as the malevolent and the virtuous, if we are to keep our commonsense moral practice from being nothing more than a conceit. Through that practice we encourage, implore, entreat, deter, dissuade, forbid, and, ultimately, praise and blame each other's attitudes, choices, decisions, and actions. But such interpersonal transactions would be, at best, benighted and, at worst, fraudulent if there is nothing any of us can do about these matters. The integrity of our moral practice demands that it be possible simply not to care about what is morally significant. If it is not possible to be morally indifferent we cannot help but wonder just what is so noble and praiseworthy about being morally good. What is so special about having a favorable attitude towards virtuous behavior if such an attitude is necessitated by simply making the judgment that such behavior *is* virtuous? How hard is it, really, to judge that keeping promises is right or that ignoring the neediest is wrong? Isn't the trick, after all, to get people to *care* about such things? What is so pressing if they care by necessity? By claiming a necessary connection between moral judgment and motivation internalism seems only to obscure the dynamic nature of our moral practice, it does nothing to illuminate it. What is truly compelling about the virtuous, no less than the vicious, appears to be beyond the internalist's scope. Hence it is surprising to find a weak internalist such as Smith attempting to trump the externalist's argument from amoralism with an argument to the effect that weak internalism, as opposed to externalism, can best account for the motivations of the "good and strong-willed" (MP: 71).[50] Is it possible that weak internalists

---

[50] This argument appears in a number of places, most notably *The Moral Problem*, 71-6; "The Argument for Internalism: Reply to Miller" *op. cit.*, and "In Defense of *The Moral Problem*: Reply to Brink, Copp, and Sayre-McCord" *op. cit.*, 111-17. In "Reply to Miller" and "In Defense of *The Moral Problem*" Smith uses the term

can explain moral virtue in a more intuitively satisfying manner than externalists, so much more so that we would feel obliged to join the internalists in denying the existence of the rational amoralist? I want to conclude this chapter by arguing that they cannot.

The argument for internalism begins by with the observation of a "striking fact about moral motivation," namely "that a *change in motivation* follows reliably in the wake of a *change in moral judgment*, at least in the good and strong-willed person" (MP: 71, emphasis in the original). This "striking fact" should be familiar, for it is simply the compliment of the metaethical requirement MR that we encountered earlier. It must be the case that morally good people are motivated in accordance with their moral judgments—and if those judgments change so, too, do the motivations—if it is a conceptual truth that a person who is *not* motivated in accordance with such judgments is immoral,[51] as MR claims. In calling MR a 'metaethical requirement,' we meant that any metaethical theory must 'respect' this principle on pain of not being a meta*ethical* theory at all. MR—as well as Smith's "striking fact"—is a moral *datum*, and data are explained, not denied. We should then agree with Smith when he says that "a plausible theory of moral judgment must therefore explain this striking fact" (MP: 71). What is at issue though is how various metaethical positions fare at the endeavor. Smith's assessment of matters is unequivocal: "As I see it, those who accept the practicality requirement can, whereas strong externalists[52] cannot, explain this striking fact in a plausible way" (MP: 71). The argument that Smith presents against the externalist's explanation of the reliability of moral motivation in morally good agents takes the form of a

---

'moralist' in favor of the phrase 'good and strong-willed' used in *The Moral Problem*. This change is purely terminological, not substantive. I will frequently combine the two and speak of the 'morally good' agent.

[51] With the necessary qualification, of course, that the inappropriate motivation is such that the agent can be held responsible for it.

[52] A 'strong' externalist is Smith's label for those who deny the practicality requirement (PR). See pg. 63 of *The Moral Problem*.

reductio ad absurdam. The externalist model of the morally good agent's motivation entails a distorted conception of the morally good agent. The internalist model does not; on this picture morally good agents appear precisely as we expect. Hence the internalist model is to be preferred. In order to assess this argument we need to be clear about what these models involve and the conception of the morally good agent that they present.

Externalists hold that the link between motivation and moral judgment is a contingent one. In other words, they deny that there is anything about the content of a moral judgment that necessitates motivation and, therefore, any motivation an agent may have with respect to some moral judgment must have its source in something *external* to that judgment. This is what contrasts them with internalists, who hold that the source of motivation is *internal* to the content of a moral judgment. As Smith says, "it follows directly from the content of moral judgment itself," which is to say "that the belief that an act is right *produces* a corresponding motivation" (MP: 72, emphasis in the original). And this difference puts the internalist, Smith claims, at a decided advantage over the externalist in explaining the striking feature of the morally good agent's motivations. Consider the following example. Say an agent believes that keeping her promise to her spouse to quit smoking is right and, further, is motivated to keep that promise. There are two ways of explaining this agent's motivation, Smith claims, one that externalists must give and the one offered by internalists. What the externalist must say is that the motivation to keep the promise is *derivative* from the general motivation to do what is believed to be right *as such* (what Smith calls the motivation to do what is believed to be right, *de dicto*). That is, the agent is motivated to do *whatever* she believes is right, so long as it is right. This motivation to do what is right, combined with a belief that a particular course of action is right, can generate a new, derived motivation to perform the particular action. The internalist's explanation, on the other hand, takes the

belief that keeping the promise is right to be sufficient of itself to produce a non-derivative, or direct, motivation to, in this instance, keep the promise (an example of what Smith calls the motivation to do what is believed to be right, *de re*). No appeal is needed here to a general concern with rightness as such, a motivation that is external to the moral judgment. But what is it about this general motivation to do what is right that saddles the externalist with an unacceptable conception of the morally good agent?

Smith brings the problem for the externalist into relief by drawing an analogy with an argument that Bernard Williams has made against moral theories that stress impartiality.[53] Williams argues that moral theories that stress impartiality serve to alienate agents from what should be most morally significant to them, a point nicely expressed in his example of a man on a beach who sees two people in the water drowning, one a stranger and one his wife. The man, of course, is in a position such that he can possibly save only one. If, in determining what he morally ought to do, the man were to employ a moral theory stressing impartiality, his choice to save his wife could only be justified if the theory sanctioned something along the lines of 'in situations of this kind it is at least all right (morally permissible) to save one's wife'. But, Williams objects,

> this construction provides the agent with one thought too many: it might have been hoped by some (for instance his wife) that his motivating thought, fully spelled out, would be the thought that it was his wife, not that it was his wife and that in situations of this kind it is permissible to save one's wife.[54]

Similarly, Smith contends, the externalist's account of the moral motivation of the morally good agent, which makes essential reference to the general, or *de dicto* motivation, to do what

---

[53] Bernard Williams, "Persons, Character, and Morality," *Moral Luck* (Cambridge: Cambridge University Press, 1981): 1-19.

[54] Ibid., 19.

is right, is equally unsatisfying, leaving good agents burdened with one thought too many. And this, he adds, is not at all how we ordinarily conceive of such agents:

> For commonsense tells us that if good people judge it right to be honest, or right to care for their children and friends and fellows, or right for people to get what they deserve, then they care non-derivatively about these things. Good people care non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve, justice, equality, and the like, not just one thing: doing what they believe to be right, where this is read *de dicto* and not *de re*. Indeed, commonsense tells us that being so motivated is a fetish or a moral vice, not the one and only moral virtue (MP: 75).

But this is unacceptable. We wanted an account of the reliability of the motivation of morally good agents, not an account of such agents that is so offensive to commonsense. Hence Smith draws his conclusion:

> We have therefore found a decisive reason to reject the strong externalist's explanation of the reliable connection between moral judgment and the motivation in the good and strong-willed person. For, in short, the strong externalist's explanation commits us to false views about the content of the good person's motivations: it elevates a moral fetish into the one and only moral virtue. And the remedy, of course, is to retreat to the alternative, internalist, explanation of the reliability of the connection between moral judgment and motivation (MP: 76).

The externalist's argument from amoralism is trumped by the internalist's argument from moral virtue. Hence, there really are no rational, yet morally indifferent agents.

Externalist rejoinders to this argument have been largely defensive in posture, primarily consisting in attempts to show that it is either not necessary to appeal to the general motivation to do what is believed to be right to explain the reliability of the morally good agent's motivation or that such an appeal does not invariably describe someone with a 'moral fetish.'[55] Though I am somewhat sympathetic with the externalist's desire to resist the

---

[55] See, for instance, Hallvard Lillehammer, "Smith on Moral Fetishism," *Analysis* 57 (1997): 187-95; David Brink, "Moral Motivation," *Ethics* 108 (1997): 4-32; Sigrún Svavarsdóttir's "Moral Cognitivism and Motivation," *Philosophical Review* 108 (1999): 161-219; and James Dreier, "Dispositions and Fetishes: Externalist Models of Moral Motivation," *Philosophy and Phenomenological Research* 61(2000): 619-618; and Jonas Olson, "Are Desires *De Dicto* Fetishistic?" *Inquiry* Vol. 45, No. 1 (2002): 89-96.

charge of 'fetishism,' this is not the tack I wish to pursue here. The externalist, I believe, is best served not by exposing weaknesses in this argument for weak internalism but in the weak internalist position itself. I am willing to grant, then, that the externalist is committed to explaining the reliability of the morally good agent's motivation by appeal to the general motivation to do what is right. Now if that renders the morally good agent a fetishist, so be it. What I want to argue, however, is that the weak internalist is, at best, in precisely the same boat, and, at worst, guilty of undermining an essential presupposition of our moral practice, namely that moral agents are responsible and therefore worthy of praise and blame.

Why say that the weak internalist is in the same boat as the externalist in explaining the morally good agent's motivation? Consider just what the role of the general motivation to do what one believes is right is meant to fill, according to the externalist. This general motivation serves as a (not necessarily conscious) conceptual bridge between the belief that something is morally right and the motivation to do it. If that bridge is not there—and it need not be, it is contingent after all—an agent may not be motivated at all to do what she judges is morally right.[56] It is the presence of this motivation to do the morally right thing that is constitutive of the morally good agent, and what guarantees the reliability of her motivations. Now whether this is simply an offensive caricature of the morally good agent or not, it seems to me to be exactly what a weak internalist like Smith is committed to himself. According to his own account of moral judgment that we encountered earlier, an agent's judgment that some course of action is morally right just is her judgment that she would be motivated to perform that action if she were fully rational. Now, putting aside the intrinsic implausibility of this account, why isn't it the case that what makes an moral agent a morally *good* agent is that she has a motivation to do what she judges she would be motivated

[56] See Sigrún Svavarsdóttir's "Moral Cognitivism and Motivation," 201.

to if she were fully rational, *whatever that might happen to be*? Why doesn't this motivation serve, in exactly the same way as the externalist's motivation to do what one believes to be right, as the conceptual bridge between particular moral judgments and motivation? Indeed, shouldn't we *expect* precisely this, given that the judgment of what one would be motivated to do if one were fully rational is being offered as an account of the *content* of moral judgment? For isn't that, after all, the fundamentally distinct feature the weak internalist is introducing into the metaethical debate?

Externalists claim that to be reliably motivated to do what one judges to be morally right one needs to have the more general motivation to do whatever it is that one judges to be morally right. The weak internalist, *qua* Smith, claims that to judge that something is morally right is properly understood in terms of judging what one would be motivated to do if one were fully rational. Now how does this claim, if true, obviate the need to have the general motivation to do whatever it is that one judges to be morally right in order to assure that an agent is reliably motivated to do what she judges to be morally right? There is no reason to think that it does. The felt need that the externalist responds to by positing *de dicto* moral motivation would, if it was ever really there to begin with, *still* be there if moral judgments are understood along weak internalist lines. Nor is there anything in the weak internalist account that would suggest that that need was illusory, the byproduct of a mistaken conception of moral judgment. Nor does the 'irrationality' of those not motivated to do what they judge to be morally right signal anything of significance in this respect. For the irrationality of the morally unmotivated is not some salient feature of wrongdoing that the weak internalist account of moral judgment serves to illuminate properly but is rather nothing more than an artifact of that idiosyncratic account. On the question of how the reliability of the morally good agent's motivation is to be explained, there seems no

detectable difference between the externalist and weak internalist accounts. If either of them paints an ugly picture of good people, then they both do.

Whether I am correct or not that the weak internalist is committed to the same *de dicto* construal of the morally good agent's motivation as the externalist, such a theorist ought nevertheless to embrace it. The reason for this is that to insist, along with Smith, that the weak internalist is properly understood as attributing to the morally good agent *de re* moral motivation is to make it difficult to see just why the morally good agent deserves any praise at all. If the morally good agent's motivation "follows directly from the content of the moral judgment itself," without the help of the conceptual bridge that *de dicto* motivation offers, then just as weak internalists maintain, such an agent is necessarily motivated in accordance with her moral judgment. But we now face what amounts to the very same question we were confronted with when considering the weak internalist claim that the immoral agent is necessarily irrational. The question facing us then was why should we count an agent who fails to be motivated in accordance with her moral judgments as responsible, and therefore blameworthy, when her failure to be motivated is due to her being irrational? Intuitively, we do *not* blame such people for their lack of motivation, and if they are not blameworthy they are not immoral. The question we face now is why we should count an agent who is motivated to do what she judges to be morally right as responsible, and therefore praiseworthy, when her motivation is necessitated by her very judgment that it is right? Intuitively, we do *not* praise people for the motivations they cannot help but have, and if they are not praiseworthy they are not morally good. Making moral motivation a necessary product of moral judgment appears to take all the *worth* out of it: it requires no special effort on the part of the agent.

The point I am pressing here is precisely the complement to that made in the earlier discussion of irrationality: without conative autonomy we lack one of the essential prerequisites that make us appropriate objects of moral address. If moral judgments necessitate motivation solely in virtue of their content then conative autonomy is superfluous and any *thing* capable of having states with evaluative content could be expected to be motivated appropriately. But this does not strike me as particularly attractive. Indeed, I think that if this account of matters is true then our moral practice of evaluating agents with respect to, and in virtue of, their motives is thoroughly misguided. On such an account there is nothing left to praise or blame. Nor would there be any need for various moral distinctions we make, such as between the good, the bad, and the indifferent: there would only be the morally good and the irrational. This thoroughly revisionist conception of the moral landscape is precisely what the weak internalist offers, as Smith himself emphatically states:

> internalists think that the division of agents into the class of moralists [the morally good] and amoralists [the morally indifferent] marks nothing in reality: it marks nothing in reality because *everyone capable of making a moral judgment is a moralist* (RM: 179, bracketed text and emphasis added).

The rather startling upshot of this is that all the non-morally good people out there—and there are plenty—are simply incapable of making a moral judgment. But if this is what the weak internalist position amounts to then it, and not externalism, is in danger of being reduced to absurdity. The conclusion we should draw is clear. Weak internalism is simply an untenable position in metaethics: cognitivists ought to be externalists.

4

Judging Wrong(ly)

> I am sorry then, I have pretended to be a philosopher: For I find your questions very perplexing; and am in danger, if my answer be too rigid and severe, of passing for a scholastic; if it be too easy and free, of being taken for a preacher of vice and immorality. However, to satisfy you, I shall deliver my opinion upon the matter, and shall only desire you to esteem it of as little consequence as I do myself. By that means you will neither think it worthy of your ridicule nor your anger.
>
> —David Hume, "The Sceptic"[1]

> Good basketball players don't think; they react.
>
> —Lou Carnesseca

Summary of the argument so far

In Chapter One we considered the famous Socratic claim that all wrongdoing is a matter of ignorance. We found that what makes the Socratic claim so counterintuitive as a general claim about wrongdoing is that our conception of wrongdoing is essentially conative/affective in nature. The categories of immorality, such as wickedness, weakness, indifference, and negligence all involve, essentially, morally problematic motivation. Indeed, such categories are really labels for certain morally problematic motivational profiles. For it is precisely an agent's motivation that renders her immoral and leads her to act (and react) in objectionable ways. We also found that our moral explanations of immoral agents and their immoral behavior make ineliminable reference to their motivation or motivational capacities. Only by appealing to such things as an agent's desires and attitudes, insufficient motivation, or lack of concern altogether for the moral significance of one's actions that we can hope to account for behavior that can be blamed.[2] To say that wrongdoing is due to the mistaken

---

[1] David Hume, *Essays: Moral, Political, and Literary* (Indianapolis: Liberty Fund, 1985, Revised Edition), 161-2.

[2] Ronald Milo, in his book *Immorality*, *op. cit.*, 234, identifies the three primary causes of immorality as bad preferences, lack of control (of ones desires, emotions, etc.), and lack of moral concern, conative phenomena

moral judgments of the wrongdoer, to some cognitive, as opposed to conative, defect cannot be correct. Such mistakes—which is what they are if they are not the result of a condemnable motivation—are not to be blamed but pitied and lamented.[3]

Accepting a conative analysis of immorality led us to conclude that a certain account of moral judgment, one where such judgments are understood to be intrinsically motivational, could not be correct. If moral judgments are by their very nature motivational then the problematic motivation indicative of immorality can only be explained in terms of problematic judgment. That moral judgments are essentially motivational was the view held by Socrates and it led him, consistently, to the ignorance thesis. It was the argument of Chapter Two that the essential elements of that account of moral judgment are held by certain contemporary moral philosophers and that they, like Socrates, are led inevitably to misconstrue the conative nature of immorality. They, too, are forced to conceive of immorality as a consequence of mistaken moral judgment. But the various ways of wrongdoing presuppose that judgment (moral or otherwise) and motivation can diverge. The motivational nature of immorality precludes the 'strong internalist' account of moral judgment: moral judgments are not themselves motivational.

---

all. At a more general level of explanation we might appeal to such aspects of the agent's personality or character such as her cruelty, deceitfulness, greed, envy and the like. Such appeals, of course, will ultimately be resolved into appeals to the agent's preferences and concerns, in other words her conative states.

[3] The general tenor of this argument is nicely expressed in an argument of Hume's against moral rationalism (cognitivism):

> These false judgments may be thought to affect the passions and actions, which are connected with them, and may be said to render them unreasonable, in a figurative and improper way of speaking. But tho' this be acknowledg'd, 'tis easy to observe, that these errors are so far from being the source of all immorality, that they are commonly very innocent, and draw no manner of guilt upon the person who is unfortunate as to fall into them. They extent not beyond a mistake of *fact*, which moralists have not generally suppos'd criminal as being perfectly involuntary. I am more to be lamented than blam'd, if I am mistaken with regard to the influence of objects in producing pain or pleasure, or if I know not the proper means of satisfying my desires. No one can ever regard such errors as a defect in my moral character.

A Treatise of Human Nature, III, 1, i (459-60).

Further reflection on the need to keep conative states and capacities autonomous from cognitive ones led us to conclude in the third chapter that the proper analysis of moral judgments is not one that views them as relationally motivational. 'Weak internalists' maintain that though cognitive judgment and motivation are ontologically distinct they are nevertheless related by psychological necessity. According to this view, in a normal, or properly functioning agent (often expressed as being 'rational'), making a moral judgment necessarily produces, or leads to, the appropriate motivation. But as we saw in Chapter Three, even this sense of necessary connection is too strong. The only way, according to the weak internalist, for the connection between judgment and motivation to be defeated is for the agent to suffer from some sort of psychological dysfunction or disorder—something that disrupts the normal connection or process which forms that connection. But this is no more consistent with the concept of immorality than the strong internalist proposal.[4] Psychological dysfunction is not something we usually hold agents responsible for, hence the problematic motivation that such dysfunction induces is not something we generally blame an agent for having. For instance, an agent who is indifferent with respect to her moral judgments *as a result of some psychological dysfunction* does not, according to our common practice, count as immoral. To qualify as immoral we feel it is necessary that an agent be able to autonomously determine her motivation with respect to her judgments of what is moral (or rational). If such motivation is determined by what amounts to brute psychological laws relating moral judgments and motivation then our commonsense conception of moral worth—both positive and negative—would need to be revised if not replaced. The conative nature of immorality, as we commonly understand that notion,

[4] It is not that commonsense does not acknowledge such disorders but rather that it does not acknowledge them to be the proper analysis of the general phenomenon of moral wrongdoing.

demands that cognition and conation be both independent and independently evaluable. Hence we seem forced to conclude that the truth of cognitivism about moral judgment necessitates the falsity of internalism about motivation. This summarizes the argument so far.

The primary concern of this essay, however, is whether moral cognitivism itself—the moral psychology underlying moral realism—is a psychology adequate to our concept of immorality. That is, can *cognitivism* adequately allow for and satisfactorily explain the diverse ways of (obviously culpable) wrongdoing? To that end, the preceding discussion of motivational internalism has been largely preliminary. It has, however, served two important purposes. First, it has repeatedly emphasized the conative nature of immoral phenomena: what renders an agent immoral (as well as her subsequent behavior) is her motivational profile; it is her motivations that are the fundamental object of condemnation. Second, the discussion has enabled us to see that the only way that cognitivism might plausibly account for our commonsense practice of judging agents, their attitudes and their actions immoral is if it is unencumbered by the adjunct motivational thesis of internalism. What this means is that if cognitivism so understood—that is, externalist cognitivism—*still* cannot adequately allow for and explain immorality then the problem must be in the cognitivist perspective itself. If this is correct then cognitivists, be they internalists *or* externalists, adhere to a theoretically problematic thesis about the psychology of moral judgment, a thesis that is both fundamentally out of step with our commonsense practice of judging agents to be immoral and, more importantly, incapable of explaining that practice. In this chapter I will argue for these claims. Cognitivism is a deeply unsatisfactory moral psychology. If our commonsense practice of judging agents to be immoral is to be

preserved and illuminated then we cannot see such judgments as being fundamentally cognitive in nature.

Cognitivism about moral judgment, as we defined that thesis in Chapter One, is, rather simply, the psychological thesis that moral judgments are essentially representations of certain things or phenomena (actions, motives, intentions, people, states of affairs, etc.) as having moral value. Cognitivism construes moral judgment as a process that depicts, or describes moral value as a 'real' feature of those things that possess it.[5] Now what we have so far argued is that if someone who, say, executed a mentally retarded person for a capital crime were to be blamed for so acting, if they were to be found to be acting immorally in so acting, it would be in virtue of their motivation to so act. Hence, for a cognitivist, the judgment that the executioner is immoral would be the judgment that her *motivation* to so act has some feature, or property, which serves to make such motivation immoral. Put baldly, cognitivism claims that *what* makes the executioner's (or anyone else's) motivation immoral is something that we can (and do when we judge correctly) *cognize*. In other words, we can describe, depict, and, ultimately, represent in propositional form those features of the motivation that render it immoral. The primary argument of this chapter is that human moral evaluators can do no such thing and that we can have no hope of a comprehensive understanding of immorality if we persist in thinking otherwise. Moral cognitivism—and the realist metaphysic it presupposes—should be rejected.

My argument for this conclusion is different in kind from those who seek to undermine cognitivism or, more directly moral realism, by arguing that there *aren't* any morally

---

[5] That such features are taken to be 'real' (and, in turn, 'objective') is precisely what makes cognitivism the moral psychology of choice for moral realism. The operative notion of 'real' (or 'objective') is, according to David Brink, a leading contemporary moral realist, that moral 'facts' or 'truths' are independent of the evidence we have for them. See his *Moral Realism and the Foundations of Ethics*, *op. cit.*, 17-ff. This understanding of the realist nature of moral facts or properties is neutral about the nature of such properties (are they natural or non-natural?).

significant properties for evaluators to cognize, given no properties of such a mysterious, or 'queer' nature exist.[6] Nor is it akin to the argument that moral facts are unnecessary for the purposes of moral explanation, and therefore we have little reason to think that they *must* exist.[7] Indeed, I believe that moral significance—and in particular, immorality—are comprehensible features of our experience but they are not of a form to which we have cognitive access, nor do they have their significance independently of our manner of acquaintance with them. Those features that render motivations morally evaluable are not susceptible to depiction, description, or propositional representation. And, to that extent, they *are* 'queer' features of motivations.[8] My argument, rather, will be that we will have no hope of appreciating the moral significance of such features and, moreover, of understanding how they fit into the world beyond moral significance if we maintain cognitivism about moral judgments and moral realism generally. Moral cognitivism, as we might put it, is *un*philosophical.

I will present this argument in stages. In the next section I will introduce the difficulty I see with cognitivism by considering the example of the evaluation of moral judgment itself. It appears to be a fact of our moral practice that moral judgments are themselves evaluable but it seems impossible to do so within the cognitivist framework. If our practice of blaming the moral critic for her judgment is legitimate then this is a strike against cognitivist moral psychology. Just as we saw that internalist cognitivism cannot allow for various

---

[6] J. L. Mackie, *Ethics: Inventing Right and Wrong* (London: Penguin Books, 1977), is a well-known recent example of this form of argument.

[7] Gilbert Harman, *The Nature of Morality* (New York: Oxford University Press, 1977), provides an example of this form of anti-realist argument.

[8] The phrase comes from Mackie. His argument, of course, was that the truth of realism would *require* queer facts and there aren't any. I am saying that there are such queer facts but their queerness puts them beyond the scope of cognition.

forms of wrongdoing, neither internalist nor externalist cognitivism can permit immoral judgment. In the section following I will present my central argument, namely that cognitivism cannot explain the very nature of immorality (or morality). That is, if the cognitivist framework is assumed, then no answer to the question of what makes immoral motivation immoral is possible; moral cognitivism (and, ultimately, moral realism) is explanatorily vacuous. The reason for this, I will argue, is that cognitivism—and hence moral realism[9]—is committed to a non-naturalist conception of moral significance and this renders morality inexplicable. A theory that claims that its fundamental objects (moral facts) are '*sui generis*' and 'unanalyzable' is a theory that forswears all thought of explaining those objects. But a theory that cannot explain its subject matter is no theory at all. This is by no means meant as a refutation of cognitivism; it surely is not. But it is a fundamental criticism nonetheless. If my argument is correct, the truth of moral cognitivism precludes the possibility of moral theory. Our initial philosophical enterprise, that of explaining immorality, will prove impossible.

I will conclude the chapter by arguing that this claim about inexplicability enables us to see not only why cognitivism cannot account for the evaluation of moral judgments and of motivations in general, but why the appeal of internalism is as strong as it is. Internalism has proved so enticing to philosophers precisely because externalist cognitivism tells us nothing about immorality as we know it and live it outside of the study: it renders this aspect of our common moral terrain unrecognizable. In closing I will suggest where I think this leaves the philosophic enterprise of explaining wrongdoing and where the most fruitful lines of investigation are likely to be.

---

[9] The implication from cognitivism to realism is, of course, warranted only to the extent that we can preclude that moral agents are guilty of radical error. I am assuming for the sake of my argument that they are not.

Immoral moral judgment

It has been a large part of the argument so far presented that if wrongdoing were properly understood to be the result of the ignorance or misjudgment of moral agents then there really would be no such thing as immorality. For it is hardly to be disputed that in our everyday moral practice we take such mistakes to be, however unfortunate and disappointing, excusable and not blameworthy, and it is this last—being blameworthy—that we take to be an integral part of being immoral. We have employed this argument when confronted with the claim that moral judgment necessarily involves or entails an appropriate motivational orientation and so its target has so far been the internalist thesis. But the force of this argument is not limited to internalist models of motivation. This argument can be directed against cognitivism itself. For, motivational issues aside, cognitivism appears open to a very significant objection: if moral properties are features that we can (and do) cognize and if those features are themselves distinct from our cognition of them, then it would seem to follow that the very act of cognition—the making of a moral judgment—is not itself subject to moral evaluation. We either correctly judge that something has certain moral features or we incorrectly so judge, but it is not the case that our *correct judgment* is morally good or that our *incorrect judgment* is morally bad. Whether someone is virtuous or vicious, moral or immoral, would not be determined by their capacity to detect the morally significant features of things but by what they did with that information when they got it. But is this insulation of moral judgment form moral evaluation something that we should comfortably accept?

One philosopher who found this result unacceptable was Adam Smith. The version of the position that Smith found so troubling had, interestingly enough, a rather surprising

source in Francis Hutcheson. What makes Hutcheson a surprising source is that he was at the forefront of the British Sentimentalist school's critique of moral rationalism, influencing significantly Smith's own views. Nevertheless, Hutcheson's posit of a moral 'sense' that enables us to 'perceive' moral qualities has a strong cognitive flavor, and it seems to be this conception of the moral sense that led Hutcheson to the following conclusion:

> But none can apply moral attributes to the very *faculty* of perceiving moral qualities; or call his moral sense morally good or evil, any more than he calls the power of tasting, sweet or bitter; or of seeing, straight or crooked, white or black.[10]

It was precisely this conclusion that Hutcheson drew from his position that Smith thought would "be regarded by many as a sufficient confutation of it."[11] It was not, of course, that Smith disagreed with Hutcheson's premise that it is inappropriate to attribute to some faculty or sense those very properties that it was the purview of that faculty to detect, but rather that Hutcheson's argument was in fact a *reductio* of his very conception of the moral sense. Smith argues as follows:

> The qualities he [Hutcheson] allows which belong to the objects of any sense, cannot, without the greatest absurdity, be ascribed to the sense itself. Who ever thought of calling the sense of seeing black or white, the sense of hearing loud or low, or the sense of tasting sweet or bitter? And, according to him, it is equally absurd to call our moral faculties virtuous or vicious, morally good or evil. These qualities belong to the objects of those faculties, not to the faculties themselves...Yet surely if we saw any man shouting with admiration and applause at a barbarous and unmerited execution, which some insolent tyrant had ordered, we should not think we were guilty of any great absurdity in denominating this behavior vicious and morally evil in the highest degree, though it expressed nothing but depraved moral faculties, or an absurd approbation of this horrid action, as of what was noble, magnanimous, and great...There is no perversion of sentiment or affection which our heart would be more averse to enter into, or which it would reject with greater hatred and indignation than one of this kind; and so far from regarding such a constitution of mind as being merely something strange or

---

[10] Francis Hutcheson, *Illustrations Upon the Moral Sense* (1728), section 1, pg. 237, et seq.; reprinted, from the 1742 edition, in D. D. Raphael, ed., *British Moralists 1650-1800* Vol. I (Indianapolis: Hackett Publishing, 1991) section 364, pg. 311. Emphasis in the original.

[11] Adam Smith, *The Theory of Moral Sentiments* (1759), edited by D. D. Raphael and A. L. Macfie, (Indianapolis: Liberty Fund, 1984): part VII, section iii, chapter 3, paragraph 8, pg. 323.

> inconvenient, and not in any respect vicious or morally evil, we should rather consider it as the very last and most dreadful stage of moral depravity. [12]

Smith's views about what we would do in cases of the kind he describes surely cohere with our moral practice: we *do* pass moral judgment on the moral judgments of others (and, no doubt, of ourselves). If we think that, say, executing the mentally retarded is morally objectionable then we are very likely to think that that judgment itself is morally praiseworthy and that the contrary judgment is morally objectionable. When we do so, we are evaluating the *attitude*, or conative profile, the moral critic bears towards the act. We might, of course, think that someone who judges that executing the mentally retarded is morally appropriate has made an excusable mistake, but often we do not. Sometimes we condemn the person. Whether we excuse or condemn depends, greatly, on the attitude with which that person *responds* to our disagreement. Does the person wonder if she is mistaken and that her judgment may have been precipitous? Is she anxious that she might have offended? Or is she sure of herself, perhaps finding our own views to be confused, even depraved? These and other kinds or reactions seem relevant to us because we feel that the moral judgments a person makes say a great deal, morally, about her. We tend, often successfully, to predict the motivations and behavior of someone from the moral judgments they make. The moral judgments that a person makes tell us what type of person she is, whether we would want her company or eschew it. This, at any rate, is what we *do* with the information we possess about the moral judgments a person makes. We do, in fact, pass (moral) judgment on the moral judgments of others—those judgments are among our objects of evaluation. But if moral judgment takes the form that Hutcheson claimed it to

---

[12] Ibid., pg. 323.

have: if it is a cognitive process or capacity that is the essential element in such judgments, then our practice of morally evaluating such judgments becomes theoretically problematic.[13]

The problem is that if moral judgments are understood to be a cognitive process then the practice of morally evaluating moral judgments is inconsistent with our practice concerning all other cognitive judgments. In no other domain clearly involving cognition do we morally evaluate cognitive error. It is not a feature of our everyday practice to take *mistakes* of judgment—including moral judgment—to reflect on the moral character of a person. But it is only when we do understand a moral judgment to in some way express the character of a person that we feel it is an appropriate object of praise and blame. But if moral judgments are a species of cognition then it hard to see how moral judgments contrary to acceptable standards could be anything *but* a species of cognitive error; they would tell us nothing about the person's character. In that case the notion of a 'bad' moral judgment would be without moral significance, as is the notion of a 'bad' mathematical judgment. If this is so, the practice of morally evaluating moral judgments is confused. Either that or the cognitivist model of moral judgment is false.

The cognitivist's conundrum was clearly seen by the logician Arthur Prior. In his elegant *Logic and the Basis of Ethics*,[14] Prior gives the following interpretation to Smith's criticism of Hutcheson.

> At first glance it may appear that what Smith was complaining of in Hutcheson was an excess of subjectivism. But in fact it was precisely the element of objectivism or rationalism in Hutcheson which involved him in the conclusion to which Smith objected. For anyone who regards moral approbation as a perception that an object possesses a certain real character, whether this perception be the work of reason or of a sense, is bound to regard the approbation itself as beyond praise or

---

[13] We should not think that this problem for the cognitivist conception of judgment is brought on by an overly sensory-perceptual construal of such judgments. Our capacity for mathematical judgments is not subject to mathematical description.

[14] Arthur N. Prior, *Logic and the Basis of Ethics* (Oxford: Clarendon Press, 1949).

> blame. We cannot blame a man for a mistaken judgment (though we can blame him for not trying hard enough to arrive at a true one); of such a judgment all we can say is that it is mistaken, not that it is morally bad.[15]

When moral judgment is construed as a form of moral detection the evaluation of moral judgment itself is made impossible (at any rate, unjustifiable). If this is the correct construal of moral judgment then we ought to stop holding people morally accountable for the moral judgments they make. But such a revision in our moral practice would be extraordinary and should only be adopted in the absence of overwhelming counter-argument. We need to ask ourselves if we really think that that practice is fundamentally misguided. Which claim do we find more counterintuitive, that moral judgment is not cognitive in nature, or that moral judgments tell us nothing about the moral character of the people who make them? If the latter, then the rejection of cognitivism is our only recourse. Prior continues:

> It is only if we regard approbation, not as a judgment about an emotional response, but as an emotional response to another emotional response, that we can think of it as itself a possible object of approbation: for this last would then merely mean, thinking of it as itself possibly evoking an emotional response.[16]

The very evaluability of moral judgments is a function of their nature: if they are construed cognitively then they are not themselves objects of evaluation. And if we insist on the evaluative nature of moral judgments then we must construe them *non*-cognitively, something along the lines of an 'emotional response'.[17]

If moral judgments are to be evaluable then they must be understood as conative responses to certain phenomena, not descriptions or representations of such phenomena. Only if moral judgments are fundamentally conative can we extend evaluation to those very

---

[15] Ibid. 73. Compare the anti-rationalist argument of Hume cited in n. 2 above.

[16] Ibid. 73.

[17] It is worth noting that Prior himself was a cognitivist about moral judgment. See his introduction to *Logic and the Basis of Ethics*.

judgments. We can do this because there is no difficulty in conceiving of a continuous chain of conative responses: each new response can be itself responded to. You are annoyed, for instance, by my desire to smoke: I resent your being annoyed; you are angered by my resentment: I am angered by your anger; Tom is bemused by the entire exchange while Sally is increasingly anxious and frightened as the emotional temperature rises. Arguably the primary objects towards which our conative responses are directed *are* the conative responses of others (and our own). But though we can have similar iterations of cognitive states—we can have beliefs about our beliefs, and beliefs about our beliefs about our beliefs, etc.—it not possible to have evaluative beliefs about evaluative beliefs, at least not evaluative beliefs that imply the conferring of praise or blame. The very idea is a conceptual confusion.

Cognitivism and the explanation of moral significance

If the arguments of Hutcheson, Smith, and Prior are correct, cognitivism about moral judgment cannot allow for a certain focus of immorality, namely immoral moral judgment. This in itself does not imply the falsity of cognitivism for it is possible for the cognitivist to maintain that our pre-theoretic moral practice, which includes the moral evaluation of moral judgments themselves, is confused. Our moral practice of evaluating the moral judgments of others (and ourselves) may itself be without moral justification. The committed cognitivist might claim that what the arguments of the previous section perhaps suggest is that we ought to revise our moral practice in the light of this theoretical consequence of cognitivism. I believe, however, that any such revisionary approach should be eschewed, and for the considerations of methodological conservatism outlined in Chapter One. Nevertheless, I do not wish to join this issue here. Rather, I want to focus on what I take to

be the fundamental theoretical difficulty of cognitivism, namely that it has no explanatory power with respect to moral significance: immorality (all of morality, really) would be inexplicable if cognitivism were true. Indeed, this consequence of the cognitivist framework—including the metaphysics of realism—threatens the very cogency of a theory *of* morality (immorality included).

It is worth stating at the outset what this claim is *not*. Gilbert Harman has famously argued that it is entirely plausible that moral facts play no role whatsoever in the explanation of our moral judgments.[18] Our judgment that executing the mentally retarded for capital crimes is wrong, for example, most likely is *not* explained by the *fact* that executing the mentally retarded for capital crimes is wrong. This contrasts with most other areas of knowledge—particularly natural science—Harman contends, where we *do* think that our judgment that, say, the earth has only one moon is best explained by the fact that the earth *does* have only one moon. Whatever we might make of Harman's argument, it is not my concern here. The claim that I am making is that the cognitivist/realist (hereafter simply cognitivist) framework cannot explain the *wrongness* of executing the mentally retarded for capital crimes, particularly the very immorality of an agent who is motivated to so act. What cognitivism cannot provide is an answer to any generic version of W. D. Ross' question "What makes right acts right?".[19]

I say that cognitivism cannot provide an answer to the 'generic' version of Ross' question because an answer to the more specific question is of course possible. Indeed, as Ross himself answered it, what makes right acts right is their being "'morally suitable' to the

---

[18] *The Nature of Morality*, *op. cit.*, Ch. 1.

[19] This question was the title of the second chapter of Ross' *The Right and the Good* (Oxford: Oxford University Press, 1930).

situation in which the agent finds himself."[20] But this does not get us very far, as Ross himself conceded. We are still left wanting an answer to the question of what makes morally suitable actions morally suitable? What, that is, is the nature, or essence, of moral suitability (or unsuitability)? Is it that such actions are obligatory? Optimific? Where we are ultimately driven is to the question "What is the nature of moral significance?" But if cognitivism is correct no answer can be given. Moral significance—moral value—*has* no nature: it is non-natural.

You will recall that this essay opened with G. E. Moore's claim that "What is good?" and "What is bad?" are the fundamental questions for ethical theory. We immediately urged that the question "Why do we do wrong?" is also question of considerable theoretical interest and of paramount practical significance. Its theoretical interest derives from the fact that any ethical theory is necessarily constrained to allow for and explain wrongdoing. For without wrongdoing, ethical theory would seem hardly worth the bother. Its practical significance is a function of our desire that our lives be better than they are and the belief that they could be so if we were to minimize the wrongs that are done. Wrongdoing—in particular, immorality—is the fundamental moral datum, that which gives ethical theory its import. The better part of this essay has been spent arguing that the cognitivist framework is inadequate to the task of allowing for this datum. This is because the moral psychology implicit in that framework in incompatible with the very concept of blameworthy wrongdoing. But if our investigation of the adequacy of a theoretical framework in ethics is to be at all thorough we ought to consider how it fares on the most fundamental theoretical questions. Let us focus here, then, on the question "What is immorality?" The very existence of moral philosophy rests on their being such a thing, even if only in theory and

---

[20] W. D. Ross, *Foundations of Ethics*, *op. cit.*, pg. 146.

not in practice. The argument here is that the cognitivist framework cannot answer it. Immorality is necessarily non-natural, and therefore inexplicable, according to cognitivism.

I need to be clear on precisely what I mean by saying that immorality is non-natural according to cognitivism. Being so will further indicate what my argument is not. The classical understanding of a non-natural property—ethical or otherwise—is that of Plato's Theory of Forms. A Form, for instance, the Form of Immorality, is a non-natural object in the sense that it is ontologically distinct from natural objects, objects of which we can have empirical knowledge. Such objects have always been theoretically problematic. Questions about just how we might have knowledge of such objects, if not by way of our senses, immediately beg to be asked. The answers that have tended to be given have often met with incredulity. Plato himself felt the need to posit the pre-existence of the soul at which time it has access to the Forms, with the knowledge thereby gained being recollected by the soul while serving its earthly penance. Nor have philosophers of a naturalist persuasion found subsequent accounts appealing to some 'moral sense' or 'faculty' of intuition that provides direct, non-inferential knowledge of moral value any more to their liking.

Perhaps the most direct objection to such ontologically-other entities has been given by J. L. Mackie. Mackie's complaint is that describing something like immorality as a non-natural property is to render it 'queer' and mysterious. Such an entity would be unlike anything else that we know of in nature. This intrinsic oddness alone should make us wary of any claims to its existence. But Mackie goes further, questioning how such a strange and utterly distinct object could be related to those natural objects to which it is to be ascribed. Mackie puts the objection this way:

> Another way of bringing out this queerness is to ask, about something that is supposed to have some objective moral quality, how this linked with its natural features. What is the connection between the natural fact that an action is a piece of deliberate cruelty—say, causing pain just for fun—and the moral fact that it is

> wrong? It cannot be entailment, a logical or semantic necessity. Yet it is not merely that the two features occur together. The wrongness must somehow be 'consequential' or 'supervenient'; it is wrong because it is a piece of deliberate cruelty. But just what *in the world* is signified by this 'because'?[21]

This is, I believe, a powerful argument against this kind of non-naturalism. Nevertheless, it is not the line I am pursuing here. Let me explain why.

This kind of non-naturalism, exemplified by Plato and now primarily identified with Moore and other intuitionists, has been discredited not only by moral skeptics such as Mackie, but by cognitivists as well. Most moral cognitivists are no more enamored with the prospect of being saddled with non-natural properties than skeptics and other anti-realists. Such cognitivists—ethical naturalists—have thereby taken to undermining the arguments that traditional non-naturalists have given in support of their thesis. The most well-known argument for the non-naturalist conclusion is essentially semantic in nature: from the claim that moral terms such as 'bad' or 'wrong' are *indefinable* it is concluded that the properties such terms pick out are unanalyzable, or *sui generis*. That is to say they are non-natural. This type of argument takes its most infamous form in Moore's 'open-question argument.' Simply put, the argument claims that we must take moral properties to be non-natural since no natural property can be found whose linguistic expression will serve as a synonym. Any term that we might plausibly suggest as synonymous with a moral term—and hence could function as the moral term's definition—could always be doubted. For instance, if we were to suggest that 'immoral' means the same thing as 'intentionally causing harm to others', we would be forced to admit that though it would be absurd to doubt whether what is immoral is immoral—the answer is obvious and, therefore, the question is 'closed'—it is quite conceivable that someone might doubt whether intentionally causing harm to others is

---

[21] J. L. Mackie, *Ethics: Inventing Right and Wrong*, *op. cit.*, pg. 41. Emphasis in the original.

immoral. Whether intentionally causing harm to others is immoral or not is an open question. But this, Moore claims, shows that the *property* of being immoral is not identical to the property of intentionally causing harm. Since, Moore adds, this reasoning can be applied to *any* suggested definition of immorality in natural (or, for that matter, supernatural) terms, immorality, like all moral properties, is non-natural.

The naturalist counter-argument to this view is to attack the meta-theoretical presupposition that facts about language determine facts about the world; that our metaphysics is exhausted by our analysis of language. We might, for instance, concede that moral terms are not synonymous with any non-moral terms, and so any question of the form "Is A the same as B?", where A stands for some non-moral term or expression and B for some moral term or expression, is necessarily 'open'. However, this goes nowhere towards showing that the *properties* these various terms pick out are not identical.[22] Meaning needn't, nay, shouldn't, be thought of as determining intension. Furthermore, we have examples that serve to forewarn us about jumping to hasty conclusions. For instance, before scientists discovered that genes were actually particular sequences of DNA molecules in chromosomes, the question of whether something that was such a sequence was also a gene was an open question. But this fact did not preclude the eventual discovery of their identity, and thereby providing a determinate answer which renders the question closed.[23]

There is a further argument against this type of non-naturalism, and the semantic views supporting it, put forth by the contemporary cognitive naturalist David Brink. Brink argues that if the semantic difference between moral and non-moral terms were evidence for the

---

[22] This argument can be found in a number of places. For a particularly thorough discussion of this and related issues, see David Brink, *Moral Realism and the Foundations of Ethics*, *op. cit.*, especially Ch.s 6-7.

[23] Stephen Darwall, *Philosophical Ethics* (Boulder, CO: Westview Press, 1998), 35.

non-identity of moral and natural properties, then similar semantic differences between the terms of different disciplines, for instance, biology and physics, would be evidence for the ontological distinctness of the relevant properties. Hence, by this line of reasoning, biological and physical properties—indeed, all properties from different domains—would be *sui generis* relative to one another. But this is either a *reductio* of the view that semantic difference entails metaphysical difference, or the moral cognitivist's commitment to non-natural properties is not really queer after all—all disciplines have their own proprietary subject matter, ontologically distinct from all others.[24] Either way, moral cognitivism has nothing to fear from the charge.

Nevertheless, I believe that moral cognitivism *is* committed to non-naturalism and, furthermore, such a commitment renders it implausible. I believe that there is an alternative understanding of 'non-natural' which bears this out and which is immune to the kinds of arguments just rehearsed. Consider one of the more explicit statements that Moore gives of his position on goodness.

> 'Good,' then, if we mean by it that quality which we assert to belong to a thing, when we say that the thing is good, is incapable of any definition, in the most important sense of that word. The most important sense of 'definition' is that in which a definition states what are the parts which invariable compose a certain whole: and in this sense 'good' has no definition because it is simple and has no parts. It is one of those innumerable objects of thought which are themselves incapable of definition, because they are the ultimate terms of reference to which whatever *is* capable of definition must be defined. That there must be an indefinite number of such terms is obvious, on reflection; since we cannot define anything except by an analysis, which, when carried as far as it will go, refers us to something, which is simply different from anything else, and which by that ultimate difference explains the peculiarity of the whole which we are defining: for every whole contains some parts which are common to other wholes also. There is, therefore, no intrinsic difficulty in the contention that 'good' denotes a simple and indefinable quality. There are many other instances of such qualities.[25]

---

[24] David Brink, *Moral Realism and the Foundations of Ethics*, *op. cit.*, pp. 164-5, 174-5.

[25] G. E. Moore, *Principia Ethica*, *op. cit.*, pp. 61-2.

If we are careful to avoid becoming stuck on Moore's use of the word 'definition,' we can allow ourselves to see that the sense of 'definition' he is concerned with *might* actually be that of explanation, in the sense of making something intelligible by placing it within a more or less systematic account reality.[26] Something which is explained in this sense is made understood in other terms. Such an explanation tells us what that something *is*; it provides an account of that something's *nature*. Let's call anything which can be so explained a 'natural' thing, and anything that exists but cannot be so explained a 'non-natural' thing. Now, defining these terms in this way does not involve us in any queer or metaphysical commitments. Indeed, if we like, we can say that everything that exists is natural, in the traditional sense of that word. But surely Moore is right that some of these traditionally natural objects are not such that we can provide a theory of them—an explanatory account of their *nature*. Such natural objects would then be non-natural in our new sense. How could an object of the natural world be unexplainable in this way? What Moore is trying to get at, it seems to me, is that *something* in the natural world is, necessarily, fundamental, or *ultimate*. Such ultimate features of reality, whatever they may happen to be, are such that they can be employed to explain other, less-ultimate aspects of reality, but which are themselves 'beyond' such explanation. Explanation, after all, has to come to an end somewhere.

Ultimate, and therefore non-natural features are unanalyzable; there are no more basic elements by which we can identify them. In that sense they are simple and *sui generis*. We

---

[26] Admittedly it is not obvious that Moore himself was a clear on this issue as one would like. His analogy between 'good' and 'yellow' bears this out. 'Yellow,' we may concede, is not definable in other terms, but we might plausibly think that *color* yellow can be analyzed, or explained. That is, we might think that colors can figure in a systematic account of what there is, in which case we would be able to specify the 'nature' of colors. Likewise, we may be concede that 'good' is semantically atomic without feeling obliged to think that the property of goodness cannot be explained in terms of other, more basic, features of reality.

have, of course, examples of such features, at least theoretically. There are, for instance, quanta, strings, and p-branes. These are candidates proposed as elementary particles, particles that are presumed indivisible. String-theory, for example, would claim that strings, ultimately, figure in the explanation—the theory—of everything, but if so there cannot therefore be a theory *of* strings. There is simply nothing more basic than strings that we could make reference to in an account of the nature of strings. Strings (or whatever the elementary particles in fact turn out to be) are therefore non-natural entities in our sense.

It is this sense of 'non-natural' that I suggest it is most profitable to view Moore's critique of ethical naturalism (though it is no part of my argument to suggest that this *is* what Moore intended). Goodness (and badness), according to Moore, is a fundamental aspect of reality, indivisible—and therefore unexplainable—in other terms. We needn't go farther and claim that such a property is ontologically distinct. But nor do I think that Moore's predecessors must be interpreted as making a metaphysical claim when they espoused non-naturalism. Consider, for instance, Richard Price on right and wrong:

> It is a very necessary previous observation, that our ideas of *right* and *wrong* are simple ideas, and must therefore be ascribed to some power of *immediate* perception in the human mind. He that doubts this, need only try to give definitions of them, which shall amount to more than synonymous expressions. Most of the confusion in which the question concerning the foundation of morals has been involved has proceeded from inattention to this remark. There are, undoubtedly, some actions that are *ultimately* approved, and for justifying which no reason can be assigned; as there are some ends, which are *ultimately* desired, and for choosing which no reason can be given. Were not this true; there would be an infinite progression of reasons and ends, and therefore nothing could be at all approved or desired.[27]

There seems nothing in this point about the ultimacy of rightness and wrongness that suggests that we must accept non-naturalism in its traditional, ontological interpretation.

---

[27] Richard Price, *Review of the Principal Question of Morals* (1758), chapter one, section III; reprinted, from the 1787 edition, in DD. Raphael, ed., *British Moralists 1650-1800*, section 672, pp. 141-42. Emphasis in the original.

Price's real claim is that *no reason can be given* for the rightness or wrongness of some actions. That is to say, the moral significance of such actions cannot be explained.[28]

It will be noticed that this passage from Price contains another traditional claim associated with non-naturalism, namely that our access to non-natural properties must be by a process of intuition—"an immediate perception in the human mind." This, we noted, has been a problem for non-naturalism, as many philosophers have found the idea of such access epistemologically problematic. But our explanatory sense of 'non-natural' can be used to obviate that problem. As we have seen, elementary particles in physics are non-natural in this sense but there is little reason to think that physicists would claim that our only epistemic access to them is by some faculty or process of intuition. This suggests that those moral cognitivists, like Brink, who defend the possibility of a coherentist moral epistemology, are correct in their claim that realism, particularly a non-natural realism, does not entail intuitionism.[29] The emptiness of an intuitionist epistemology is not a problem with which the moral cognitivist is necessarily burdened.

The explanatory sense of 'non-natural' commits the cognitivist neither to a mysterious metaphysics of morals nor to an embarrassingly question-begging moral epistemology.

---

[28] This is the interpretation Prior gives of Price in his account of the history of Moore's 'naturalistic fallacy.' Commenting on another passage of Price's *A Review of the Principal Questions in Morals*, Prior offers the following:

> An ethical rationalist might, of course, call an emotion or action 'reasonable' when it is 'such as our reason discerns to be right'. Price, for example, gives this as the meaning of the phrase 'conformity of our actions to reason', though he notes that as a method of 'explaining virtue' this phrase is useless—we can only talk in this way if we believe that 'rightness' and 'wrongness' cannot be 'explained' at all.

See Prior's *Logic and the Basis of Ethics*, *op. cit.*, pg. 81. Prior's interpretation of Price, it should be noted, is unaided by any explicit statement by Price that 'rightness' and 'wrongness' cannot be 'explained' at all. The relevant passage of Price states the following:

> Explaining virtue by saying, that it is the conformity of our actions to reason, is yet less proper; for this conformity signifying only, that our actions are such as our reason discerns them to be right; it will be no more than saying, that virtue is doing right.

A Review of the Principal Questions in Morals, Ch. VI, reprinted in L. A. Selby-Bigge, ed., *British Moralists*, (Oxford: Clarendon Press, 1897; republished by Bobbs-Merrill, 1964), section 694, pg. 171 of Vol. II.

[29] See Brink, *Moral Realism and the Foundations of Ethics*, *op. cit.*, Ch. 5.

Such bouquets however, significant though they may be, cannot spare cognitivism from the shroud of implausibility that even this sense of non-naturalism imposes. What our non-naturalist claims is that moral value is an ultimate aspect of reality, one which perhaps enables us to comprehend the moral dimension of our experience but which is itself too basic to be understood. The immorality of the agent who intentionally harms others for the fun of it is an elementary aspect of that agent's motivation, one that cannot be analyzed or explained. What makes that motivation immoral is something we cannot know or say since *nothing makes* that motivation immoral; rather, it simply *is* immoral.

Ethical non-naturalism, in this sense, is explanatorily vacant: there can be no explanation, no account of immorality, or any other morally significant property, if such properties are taken to be explanatorily ultimate. And this, I would suggest, is good reason for rejecting this sense of non-naturalism. There is precious little to recommend the view that moral value is such a basic element of the universe, explanatorily on a par with the elementary particles of physics. This, of course, comes nowhere near a refutation of non-naturalism, nor of any other theoretical commitment that leads to it. Nevertheless, I would daresay that many a moral cognitivist would question their adherence to that doctrine if they believed that it committed them to the ultimacy and inexplicable nature of moral value. And this, I want to suggest, is precisely the position moral cognitivists are in. If moral value is a discreet, representable feature of reality then it lacks an explainable nature. Ethical naturalism, cognitively construed, is impossible.

The impossibility of cognitive naturalism is not something for which one can offer a proof. What one can do, however, is provide reasons for thinking it to be nevertheless true. It is worthwhile, in this regard, to consider the context in which modern non-naturalism flourished. Quite interestingly, it is the reverse image of the historical dynamic of moral

philosophy in the first half of the twentieth century. Ayer's emotivism, for instance, can be seen as a direct response to the metaphysical extravagance he found in the work of the likes of Moore, Prichard, and Ross. Whereas their predecessors, such as Price and Thomas Reid, were developing their views in response to the Sentimentalism of the likes of Hutcheson, Hume, and Smith. The following dialectic between Hume and Reid proves instructive. It begins with one of the most famous passages of Hume, wherein he questions the transition from 'is' to 'ought'.

> In every system of morality, which I have hitherto met with, I have always remarked, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surprised to find, that instead of the usual copulations of propositions, *is* and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*. This change is imperceptible; but is, however, of the last consequence. For as this *ought*, or *ought not*, expresses some new relation or affirmation, it is necessary that it should be observed and explained; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it. But as authors do not commonly use this precaution, I shall presume to recommend it to the readers; and am persuaded, that this small attention would subvert all the vulgar systems of morality, and let us see, that the distinction of vice and virtue is not founded merely on the relations of objects, nor is perceived by reason.[30]

Reid, quoting this passage in full, responds as follows:

> We may here observe, that it is acknowledged, that the words *ought* and *ought not* express some relation or affirmation; but a relation or affirmation which Mr. Hume thought inexplicable, or, at least, inconsistent with his system of morals. He must, therefore, have thought, that they ought not to be used in treating of that subject. He likewise makes two demands, and, taking it for granted that they cannot be satisfied, is persuaded, that an attention to this is sufficient to subvert all the vulgar systems of morals. The *first* demand is, that *ought* and *ought not* be explained. To a man that understands English, there are surely no words that require explanation less. Are not all men taught, from their early years, that they ought not to lie, nor steal, nor swear falsely? But Mr. Hume thinks, that men never understood what these precepts mean, or rather that they are unintelligible. If this be so, I think indeed it will follow, that all the vulgar systems of morals are subverted.
>
> Dr. Johnson, in his Dictionary, explains the word *ought* to signify, being obliged by duty; and I know no better explication that can be given of it...The second

[30] David Hume, *A Treatise of Human Nature*, III, i, 1. Emphasis in the original.

> demand is, that a reason should be given why this relation should be a deduction from others which are entirely different from it. This is to demand a reason for what does not exist. The first principles of morals are not deductions. They are self-evident; and their truth, like that of other axioms, is perceived without reasoning or deduction. And moral truths that are not self-evident, are deduced, not from relations quite different from them, but from the first principles of morals.[31]

However unsatisfying this published dialogue may be, it is nonetheless telling. Hume demands that the cognitivist's conception of normativity be explained and Reid's cognitivist response is to not only refuse to meet that demand but to deny that there is any need to do so. Moral significance requires no explanation since everyone of sufficient maturity fully comprehends it anyway. The significance of the fact that the explication of *ought* that Reid in fact gives is one that is morally laden is, of course, not lost on him. But to demand otherwise is to demand "what does not exist." The crucial point for us, however, is that whether we find Reid's response plausible or not, it is essential that we do not find it unduly extreme. Broadly construed, Hume (and Smith, and Ayer, etc.) and Reid (and Price, and Moore, etc.) present us with what seem to be the only viable models for the foundations of morals. Either morality and immorality is explained in the manner of Hume, or present day expressivists, or it isn't explained at all, as Reid, Price, and Moore would have it. There does not seem to be any middle ground for the cognitivist to occupy.[32]

The logical space for explanatory theories of morality would seem to be the following. The Humean, or expressivist, offers an explanation of moral significance in naturalist terms,

---

[31] Thomas Reid, *Essays on the Active Powers of Man* (1788), Essay V, Ch. VII, reprinted in D. D. Raphael, ed., *British Moralists: 1650-1800*, Vol. II, *op. cit.*, paragraphs 928-930. Emphasis in the original.

[32] A point I take Prior to have endorsed in his introduction to *Logic and the Basis of Ethics*:

> I am inclined to think that almost all that can be said, from a purely logical point of view, on the issue between naturalism and non-naturalism, has already been said in two quite brief sections in Hume's *Treatise of Human Nature* (II. iii. 3, and III, i, 1) and in one quite brief chapter in Reid's *Essays on the Active Powers* (V. vii). At all events, a thorough mastery of these three items would provide anyone with a very complete set of tools for cutting away the thick growths of sophistry which seem in all periods to thrive on the soil of Moral Philosophy. Page x.

one where the moral value of some action or motivation is a function of the conative/affective states of the critic (or of an idealized version of the critic) contemplating that action or motivation. We can imagine more or less complex versions of such an account, but the rudiments are essentially the same: the explanation of moral value lies in the dynamic interplay among conative/affective beings such as ourselves. The immorality, say, of the motivation of an agent willfully inflicting pain on another for fun is explained, naturalistically, in terms of our conative/affective *response* to the conative/affective profile generating the action in question. Now such an explanation of moral value is decidedly what moral philosophers call 'anti-realist' and non-cognitivist, and for this reason many of those philosophers have found this kind of explanation unpalatable. Moral value, these critics claim, is a *real* feature of the universe, one that we can discover and recognize—in a word, cognize—not something that human beings *project* through their desires and attitudes on to the desires and attitudes of others (as well as other features of an essentially non-moral reality). Some of these critics, those we are calling non-naturalists, make the claim that moral value is not only real but so fundamental to the nature of the universe that an explanation of it, in any common understanding of 'explanation', is impossible. It is, they would say, a *brute* fact of reality. But in so arguing, these cognitivists have also taken other cognitivists—naturalists—to task for attempting to provide an their own explanation of morality, one that would be fundamentally different in kind from that which the non-cognitivist offers. Their argument against the naturalist cognitivist is perhaps best seen as the claim that the naturalist faces a challenge that cannot possibly met. The challenge being to provide an explanation of morality in terms of other natural—that is, explainable—features of the universe *without fundamental appeal to psychological features of moral agents and critics.*

The challenge facing cognitive naturalists, those who understand moral properties to be complex and divisible, is a daunting one. To see just how so, consider a currently popular version of moral realism, one which holds that certain non-moral properties combine in certain circumstances to constitute moral properties.[33] The structure of an account of this kind is one that we are familiar with from other domains of inquiry. For instance, clusters of chemical properties combine, in some way, to constitute biological properties. The resultant biological properties are not taken to be identical to their constitutive chemical properties, but they can nonetheless be *explained* by them. Now there seems no reason to think that this kind of explanation is in any way problematic or unacceptable; this is a legitimate way of explicating theoretical relationships. And we might, if only for the sake of the argument, allow that through and investigation into, say, our linguistic practices the realist can established a relationship between moral properties and certain natural properties by this method. But it must be admitted by the realist that such an identification comes nowhere near an explanation of just *how* some 'homeostatic property-cluster'[34] of natural properties conveys *moral significance* to the property that it 'constitutes' or that supervenes on it. Nor do we have any reason to think that an explanation of how such moral significance is effected can be produced without essential appeal to the conative/affective states and dispositions of moral agents. And it is only by respecting such a constraint on the explanation of morality that an ethical naturalist can hope to maintain the objectivity of

---

[33] This seems to be the position of the so-called 'Cornell Realists'. See, for instance, David Brink, *Moral Realism and the Foundations of Ethics*, *op. cit.*, pp. 156-63; 172-80; Richard Boyd, "How to be a Moral Realist,"in G. Sayre-McCord, ed., *Essays on Moral Realism* (Ithaca: Cornell University Press, 1988); Nicholas Sturgeon, "What Difference Does it Make Whether Moral Realism is True?" in *Moral Realism: Proceedings of the 1985 Spindel Conference*, ed. N. Gillespie, *Southern Journal of Philosophy* 4 (supp.) (1986); and Richard Miller, "Reason and Commitment in the Social Sciences," *Philosophy and Public Affairs* 8 (1979): 241-66.

[34] Boyd, "How to be a Moral Realist," ibid.

morality and, hence, the plausibility of a cognitivist moral psychology.[35] As it stands, naturalists have yet to offer any such account. The claim here, and what I take non-naturalists to believe, is that no account satisfying such a constraint is likely to be forthcoming. As far as the non-naturalists would have it, non-cognitivists have cornered the market on naturalist explanations in ethics: beyond the assessment of attitudes by other attitudes, no explicable natural properties seem plausible candidates for turning the non-moral into the moral: from turning the intentional causing of harm into immorality.

The foregoing suggests we cannot hope to have an explanation of immorality that enables us to adequately comprehend what makes immoral agents immoral if we assume a cognitivist framework for morality. And if we are to choose a theoretical framework for morality on the basis of an inference to the best explanation, the conativist or expressivist account ought to be the result. For in being the *only* approach that has an explanation of morality it is the *best* explanation by default. In any event, to maintain cognitivism and its essential non-naturalism would be unfortunate. Practically speaking, to have no substantive explanation of immorality is to preclude any constructive dealings with immoral agents. The only measure we can take with the more problematic among them would be confinement. Theoretically speaking, matters are no more satisfying. Our attempted philosophical analysis of immorality must remain, necessarily, incomplete: we will be unable to offer any theory of *what* it is about the motivation of immoral agents that renders it blameworthy. If cognitivism is true, the immorality of the indifferent, the negligent, the

---

[35] The argument of this paragraph is essentially an application of the worry Darwall, Gibbard, and Railton express concerning reductionist naturalist theories of morality to what they call 'postpositivist nonreductionist' accounts. See Stephen Darwall, Allan Gibbard, and Peter Railton, "Toward a *Fin de siècle* Ethics: Some Trends," *The Philosophical Review* 101, No. 1 (1992): 115-89, especially pp. 170-7.

weak, and the wicked's motivational profiles will be a brute, inexplicable fact and, given our rejection of internalism, a fact whose *motivational* significance will remain mysterious.

Towards a conclusion

I want to now bring the criticisms of cognitivism presented in the previous two sections together to form a diagnosis of the contemporary situation in moral psychology. We have just argued that a cognitivist moral psychology, and the realist metaphysic it presupposes, must face the challenge of 'placing' morality within the rest of our explanatory framework. It must place it either nestled among our theories of other, non-moral domains, or it must place it among the cornerstones of the theoretical edifice, something like an 'Axiom of Moral Significance.' There is a great tradition in moral cognitivism that suggests the latter. If they are correct then the 'nature' something like immorality can never be known. The appreciation of the immorality of some person's behavior is a fundamental aspect of our experience, a *given.*

There is a certain intuitive appeal to some aspects this picture, particularly with respect to the phenomenology of our moral experience. Moral value *does* seem to have the status of a 'given' of our experience; the good and the bad have been with us all our lives even when we weren't so sure just what was in fact good and bad. And the omnipresence of value breeds familiarity: it is something we feel we know intimately, perhaps too so. But for all that we can become tongue-tied when we try to say just what it is. And those who study moral value can provide no more satisfying analysis than those who reflect little on such things. Nor does there seem to be anything in nature that we can assuredly identify as being always good or bad. Nothing, that is, short of the most general and vaguest of things, like happiness or harm. But even here we see slippage. For we can often think that it is good

to deny happiness now for happiness later, or that the infliction of some harm may lead to some, or even more, good. If we did not take there to be an implicit gap between these things and their moral value we could not avoid descent into either vacuity or incoherence. Moral value, it would seem, is fundamental to the fabric of experience.

But there is also something deeply troubling about this account of moral experience. Troubling, anyway, when it is offered within a framework taking the apprehension of the moral given as a matter of cognition, a psychological process that essentially involves the representation of discreet comprehensible, entities—*values*. For it is not at all clear how something given, something so basic to the world of our experience that it resists description could possibly *be* represented. Alternatively, we might just ask how an object of cognition could be so persistently immune from rational analysis? It would seem the very obscurity of value suggests that the cognitive model is an unfortunate choice for moral psychology. Perhaps a great deal of moral philosophy rests on a mistake after all.

We argued earlier that moral judgments, which in practice are frequently appealed to in the moral assessment of people, are inappropriately so used if those judgments are essentially cognitive. The properties that a cognitive faculty is meant to discern are not attributable to the act of discernment itself. But this, too, suggests that cognitivism is out of its natural element when applied to morality. To cease evaluating people by way of evaluating their moral judgments would be an extraordinary revision in our interpersonal relations, one that we have little impetus to make. Indeed, the application of the cognitive model of judgment in morality has such counterintuitive results that it borders on incoherence; a category mistake in philosophical methodology. Moral cognitivism precludes evaluating people by their moral judgments which leaves their overt behavior or, more accurately, the *motivation* with which they behaved, as the only evaluable criterion for the purposes of agent

assessment. And it is here that we most probably will feel the intuitive pull of the internalist thesis. For though, according to cognitivism, we can learn nothing about the morality of an agent directly from their moral judgments, it might be thought that we can do so indirectly, via the necessarily intimate connection such judgments have with motivation. But this, we have seen, leads to intolerably counterintuitive results concerning the very possibility of familiar types of immorality. Ways of wrongdoing, such as amorality, negligence, weakness and, wickedness are made unrecognizable, if not impossible, when motivation is metaphysically or conceptually tied to moral judgment. In order for the motivation behind wrongdoing to confer blameworthiness on the agent it must be understood to be sufficiently autonomous from moral judgment to be evaluable on its own. But cognitivism, oddly, cannot provide any answer as to *what* it is about an agent's motivation that, for instance, renders it immoral. There are no discrete and discernable features of such a motivation that the cognitive apparatus can detect, none anyway, that confer the immorality. The immorality of a motivational orientation that counts, say, as wicked is a brute fact about it, a 'given' that we either get or don't get. But we might begin wondering at this point why the phenomenon of 'getting the given' should be thought of as cognitive at all.

In fact I think that it should not be so thought. The problem of cognitivism lies in the nature of motivation itself, the fundamental object of evaluation. It is motivation, after all, that ultimately proves to be the bane of the cognitivist perspective, making trouble for the cognitive internalists and externalist alike. According to the externalist, the internalist's commitment to a necessary link between moral judgment and motivation precludes an adequate treatment of blameworthy wrongdoing, effectively rendering the position unacceptable. According to the internalist, *denying* the essential connection between moral

judgment and motivation succeeds only in making the nature of moral judgment itself mysterious and obscure. James Lenman puts the internalist's complaint nicely.

> For the externalist holds that though someone may *believe* something good, it may nonetheless be the case that she *couldn't care less* about it. And the problem here is that this leaves it *wholly* mysterious how such a person effectively *differs* from someone who does not have the belief in question, what on earth the having of such a belief *amounts* to.[36]

It has essentially been the argument of this essay that both sides of this debate have been hitting their marks. The inability of the internalist paradigm to capture immorality, perhaps the core datum of actual moral experience and most significant impetus to theory construction in ethics, makes it unacceptable. Hence, we concluded, cognitivists ought to be externalists. But this is hardly a more advantageous position. The challenge of saying *just what* something like immorality 'amounts' to, just what the *nature* of immorality *is*, is one that no cognitivist can overcome. Again, Lenman captures the difficulty perfectly.

> Any form of moral realism invites the challenge of explaining just what the belief that something is, say, good is supposed to *amount to*, just what its *content* is. Of course, this challenge might be refused. The realist might choose to echo Moore's insistence that "good is good and that is the end of the matter". And of course, realism of this non-naturalistic kind, when sufficiently purified both of empirical content and practical significance, has the obvious merit of being irrefutable, though this comes as a package with a baffling vacuity. The realist who wishes to resist this path has no option but to meet the challenge head on, perhaps the fundamental ground for suspicion of any kind of moral realism being simply doubt that the challenge can ever be satisfactorily so met.[37]

We have argued here that this suspicion is a plausible one: no naturalist ethic satisfying cognitivist criteria is likely to be forthcoming. The non-naturalist claim that our moral judgments *have no* content beyond themselves is the only cogent claim for a cognitivist to make. 'Getting the given' is, at bottom, what moral cognitivism is all about.

---

[36] James Lenman, "The Externalist and the Amoralist," *Philosophia* 27, Nos. 3-4 (1999): 441-56, quotation from pg. 442. Emphasis in the original.

[37] Ibid., pg. 442. Emphasis in the original.

What all of this suggests, I believe, is that what really makes motivation morally significant is not something that the cognitivist paradigm of moral judgment can capture. Not because moral significance is a brute 'given' of nature, nor because moral judgment is a direct 'getting of the given', but rather because motivation—specifically those aspects of conative/affective states that give them moral significance—are simply not susceptible to the type of representation that cognition involves. To that extent, the moral significance of motivation *is* a 'queer' feature, one that our powers of reason have considerable trouble making heads or tails of. But what *can* make 'sense' of a person's motivations and attitudes are precisely *other* motivations and attitudes. If we construe moral judgment *non*-cognitively we can see possible ways out of the cognitivist's conundrum. Recall, for instance, Prior's argument that moral judgment can be subject to moral evaluation *only if* it is understood to be something along the lines of an emotional response, or an attitude, or 'stance' towards something. Only then can we make sense of *praising* and *blaming* the moral views of another; only then can we see how a person's moral views reflect on their moral character. But such a conception of moral judgment also makes available a much more reasonable—and plausible—explanation as to why we cannot articulate with any satisfying degree of precision just what being immoral amounts to. Whereas the cognitivist must explain this by claiming that immorality is a fundamental and ultimate aspect of experience, the non-cognitivist need only state the far more modest claim that if moral judgments are non-cognitive in nature than it follows that such 'judgments' have no analyzable content.

Construing moral judgments along the lines of motivational states has the further advantage of bringing a natural symmetry to the moral realm. If motivations are the proper *objects* of evaluation it seems appropriate to have them evaluated by other motivations; to view the moral life in terms of attitudes assessing attitudes. Equating moral judgments with

motivational states does of course face considerable objections, not the least of which is that such a view seems hopelessly internalist, making it subject to the very same objections we leveled at the cognitivist variant of that view. But I believe that a plausible defense of the conative account of judgment can be mounted, one that trades on the *lack* of an essential cognitive aspect to judgment. Indeed, I believe that the cognitive externalist response to the problems besetting the cognitive internalist gets things the wrong way round. The problem is not taking moral judgment to have, necessarily, a motivational aspect, but rather taking it to have a cognitive one.

I will not build a defense for this claim here, but I will close with a thought about what a conativist perspective would be like. Conativism is generally thought to necessarily deny the authenticity of certain ways of wrongdoing, allowing an agent to do what is judged to be wrong only if that judgment is insincere, one made in an 'inverted-commas' sense.[38] The conativist, it is claimed, is necessarily skeptical about knowingly willful wrongdoing. And this shows poor philosophical taste. As David Brink puts it:

> But, again, skeptical solutions are in general not very satisfying responses to skeptical problems; they dispose of disturbing challenges too easily. The noncognitivist's victory over the amoralist also seems to be won too easily. We can imagine someone who sincerely thinks that some action, say, is wrong—and not simply that it is conventionally regarded as wrong—and yet remains unmoved. If so, the amoralist is conceivable and we should not simply dismiss amoral skepticism as incoherent. At the very least, we should accept a skeptical solution to this skeptical problem only when all straight solutions have failed.[39]

This criticism is unfair, and the conativist need accept no such description of her account. The amoralist, the person who sees that some action would bring misery, deny autonomy, or would break a promise and yet feels no compunction about performing it, is not confused,

---

[38] See for instance, R. M. Hare, *The Language of Morals* (Oxford: Oxford University Press, 1952): 124-6, 163-5, Brink, *Moral Realism and the Foundations of Ethics*, *op. cit.*, 46-7; and Michael Smith, *The Moral Problem*, *op. cit.*, 66-ff.

[39] David Brink, *Moral Realism and the Foundations of Ethics*, *op. cit.*, 84.

she is contemptible. The conativist needn't call her incoherent but rather a monster. Likewise a person who feels a pull towards such actions—the wicked person who performs them *because* they can be so described. Such a person offends and horrifies, and to them we should be intolerant. This better captures what the conativist account of wrongdoing is than what we see through the distorting lens of the cognitivist perspective, one that makes us lose sight of what really matters in moral dispute and which puts the focus on the irrelevant. Consider, again, Brink:

> It would seem to follow from the noncognitivist semantic claims that the amoralist and moralist must be in *moral disagreement*, because they hold different attitudes towards the same actions, for example. But although there may be an important dispute between the moralist and the amoralist (i.e., the amoralist thinks moral concern irrational while the moralist does not), their dispute is not a moral one. For the genuine amoralist is precisely someone who shares the moralist's moral views yet remains unmoved.[40]

Only if we take moral judgment to amount to no more than the cataloguing and categorization of the moral dimension of experience could we possibly say that there is no moral dispute here. If you differ in attitude from me over treating people with respect or keeping your promises or abusing children or the elderly, then we most certainly have a moral disagreement. That we both engage in the linguistic practice of speaking of such actions as 'right' or 'wrong' matters not. You do not share my moral views simply (not even in part) by sharing my language, nor even my thoughts. Our moral views are a matter of what we will allow, encourage, forbid, and reject. If we do not agree on these then we are morally opposed. That conativism allows us to see immorality as an offense against what we hold dear is its greatest virtue. Cognitivism, on the other hand, must construe it merely as a failure of the mind. An appreciation of wrongdoing is something we can have; cognitivism has simply been looking in the wrong place.

---

[40] Ibid., 86. Emphasis in the original.

Bibliography

Altham, J.E.J. "The Legacy of Emotivism." *Fact, Science, and Morality: Essays on A.J. Ayer's Language, Truth, and Logic.* Edited by Graham Macdonald and Crispin Wright. Oxford: Basil Blackwell, 1986.

Anscombe, G.E.M. *Intention* . Oxford: Basil Blackwell, 1957.

—"Modern Moral Philosophy." *Philosophy* 33 (1958). Reprinted in *Virtue Ethics*. Edited by Roger Crisp and Michael Slote. Oxford: Oxford University Press, 1997.

Aristotle. *Nicomachean Ethics*. Translated by W. D. Ross. *The Basic Works of Aristotle*. Edited by Richard McKeon. New York: Random House, 1941.

Arkonovich, Steven. "Defending Desire: Scanlon's Anti-Humeanism." *Philosophy and Phenomenological Research*, LXIII (2001): 499-519.

Arpaly, Nomy. *Moral Worth*, *The Journal of Philosophy*, XCIX, (2002): 223-245.

Ayer, J. *Language, Truth, and Logic*. New York: Dover, 1952.

Bennett, Jonathan. "The Conscience of Huckleberry Finn." *Philosophy* XLIX (1974): 123-34.

Birches Thomas C., and Nicholas D. Smith. *The Philosophy of Socrates*. Boulder: Westview Press, 2000.

Blackburn, Simon. *Ruling Passions*. Oxford: Clarendon Press, 1998.

Boyd, Richard. "How to be a Moral Realist." Reprinted in *Moral Discourse and Practice*. Edited by Stephen Darwall, Allan Gibbard, and Peter Railton. New York and Oxford: Oxford University Press, 1997.

Brandt, Richard B. *A Theory of the Good and the Right*. Oxford: Oxford University Press, 1979. Reprinted by Prometheus Books, Amherst, New York, 1998

Brink, David. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press, 1989.

—"Moral Motivation," *Ethics* 108 (1997): 4-32.

Broadie, Sarah. *Ethics With Aristotle*. New York & Oxford: Oxford University Press, 1991.

Cooper, John M. "Some Remarks on Aristotle's Moral Psychology." Reprinted in Cooper, 1999.

—*Reason and Emotion: Essays on Ancient Moral Psychology and Ethical Theory.* Princeton: Princeton University Press, 1999.

Dancy Jonathan. *Moral Reasons.* Oxford: Blackwell, 1993.

—"Why There is Really No Such Thing as the Theory of Motivation." *Proceedings of the Aristotelian Society* (1995): 1-18.

Darwall, Stephen. "Reasons, Motives, and the Demands of Morality: An Introduction." *Moral Discourse and Practice.* Edited by Stephen Darwall, Allan Gibbard, and Peter Railton. New York and Oxford: Oxford University Press, 1997.

—*Philosophical Ethics.* Boulder: Westview Press, 1998.

Darwall, Stephen. Allan Gibbard, and Peter Railton, "Toward *Fin de siècle* Ethics: Some Trends," *The Philosophical Review* 101 (1992): 115-89.

Donald Davidson's "Actions, Reasons, and Causes," *Journal of Philosophy*, LX (1963): 685-700.

—"How is Weakness of the Will Possible." Reprinted in *Essays on Actions and Events.* Oxford, Clarendon Press, 1980.

Dennett, Daniel C. "Why You Can't Make a Computer that Feels Pain." Reprinted in *Brainstorms.* Cambridge: Bradford Books, 1978.

Dreier, James. "Dispositions and Fetishes: Externalist Models of Moral Motivation." *Philosophy and Phenomenological Research* LXI (2000): 619-618;

Dunn, Robert. *The Possibility of Weakness of Will.* Indianapolis and Cambridge: Hackett Publishing Company, 1987.

Estler Jon. *Ulysses and the Sirens: Studies in Rationality and Irrationality.* Cambridge: Cambridge University Press, 1979.

Fodor, Jerry. *Representations.* Cambridge: The MIT Press, 1981.

Frankena, W. K. "Obligation and Motivation in Recent Moral Philosophy." *Essays in Moral Philosophy.* Edited by A. I. Melden. Seattle: University of Washington Press, 1958.

Griffiths, Paul. *What Emotions Really Are.* Chicago: University of Chicago Press, 1997.

Hampton, Jean. "The Nature of Immorality." *Social Philosophy & Policy* 7. 1989. Reprinted in *The Many Faces of Evil: Historical Perspectives.* Edited by Amélie Oksenberg Rorty. London and New York: Routledge, 2001.

Harman, Gilbert. *The Nature of Morality.* New York: Oxford University Press, 1977.

Hart, H.L.A. "Negligence, *Mens Rea* and Criminal Responsibility." Reprinted in Hart, H.L.A. *Punishment and Responsibility*. Oxford: Clarendon Press, 1968.

Hill, Thomas. "Four Conceptions of Conscience." *Nomos*, LX (1998): 13-52.

Hobbes, Thomas. *Leviathan*. 1651. London: Penguin Books, 1985.

David Hume. *A Treatise of Human Nature* (1739). Edited by L. A. Selby-Bigge, 2nd Ed revised by P. H. Nidditch. Oxford: Clarendon Press, 1888/1978.

—"The Sceptic." *Essays: Moral, Political, and Literary* (1777). Rev. Edition. Indianapolis: Liberty Fund, 1985: 161-2.

Hutcheson, Francis. *Illustrations Upon the Moral Sense* (1742). Reprinted in *British Moralists 1650-1800*. Vol. I. Edited by D. D. Raphael. Indianapolis: Hackett Publishing, 1991.

Kant, Immanuel. *Foundations of the Metaphysics of Morals* (1785). Translated by Lewis White Beck. Indianapolis: Bobbs-Merrill, 1959.

Christine Korsgaard, "Skepticism about Practical Reason," *Journal of Philosophy* LXXXIII (1986): 5-26.

—"The Normativity of Instrumental Reason." *Ethics and Practical Reason*. Edited by Garrett Cullity and Berys Gaut. Oxford: Clarendon Press, 1997.

Lenman, James. "The Externalist and the Amoralist." *Philosophia* 27, Nos. 3-4 (1999): 441-56.

Lillehammer, Hallvard. "Smith on Moral Fetishism." *Analysis* 57 (1997): 187-95.

MacIntyre, Alison. "Is Akratic Action Always Irrational?" *Identity, Character, and Morality: Essays in Moral Psychology*. Edited by Owen Flanagan and Amelie Oksenberg Rorty Cambridge: MIT Press, 1990.

Mackie, J. L. *Ethics: Inventing Right and Wrong*. London: Penguin Books, 1977.

McDowell, John. "Are Moral Requirements Hypothetical Imperatives?" *Proceedings of the Aristotelian Society*, Supplementary Volume (1978): 13-29,

—"Virtue and Reason." *The Monist* 62 (1979): 331-50.

McGinn, Colin. *Ethics, Evil, and Fiction* Oxford: Clarendon Press, 1997.

McNaughton, David. *Moral Vision*. Oxford: Blackwell Publishers, 1988.

Mill, John Stuart. *Utilitarianism* (1861/63). Edited by Roger Crisp. Oxford: Oxford University Press, 1998.

Miller, Richard. 'Reason and Commitment in the Social Sciences." *Philosophy and Public Affairs* 8 (1979): 241-66.

Milo, Ronald D. *Immorality*. Princeton: Princeton University Press, 1984.

—"Virtue, Knowledge, and Wickedness." *Virtue and Vice*. Edited by Ellen Frankel Paul, Fred D. Miller, Jr., and Jeffrey Paul. Cambridge: Cambridge University Press, 1998.

Moore, G.E. *Principia Ethica* (1903). Rev. ed. Edited by Thomas Baldwin. Cambridge: Cambridge University Press, 1993.

Nagel, Thomas. *The Possibility of Altruism*. Princeton: Princeton University Press, 1970.

Nichols, Shuan. "How Psychopaths Threaten Moral Rationalism: Is it Irrational to be Immoral?" *The Monist* 85 (2002): 285-303.

Olson, Jonas. "Are Desires *De Dicto* Fetishistic?" *Inquiry* 45 (2002): 89-96.

Pears, David. *Motivated Irrationality*. Oxford, Clarendon Press, 1984.

Plato. *Protagoras*. Translated by Stanley Lombardo and Karen Bell. Indianapolis: Hackett Publishers, 1992.

Platts, Mark. *Ways of Meaning*. London: Routledge & Kegan Paul, 1979.

Price, Richard. *A Review of the Principal Questions in Morals* (1758). Reprinted, from the 1787 edition in, *British Moralists 1650-1800*. Vol. II. Edited by D. D. Raphael. Indianapolis: Hackett Publishing, 1991.

Prior, Arthur N. *Logic and the Basis of Ethics*. Oxford: Clarendon Press, 1949.

Reid, Thomas. *Essays on the Active Powers of Man* (1788). Reprinted in *British Moralists: 1650-1800*. Vol. II. Edited by D. D. Raphael. Indianapolis: Hackett Publishing, 1991.

Ross, W. D. *The Right and the Good*. Oxford: Clarendon Press, 1930.

—*Foundations of Ethics*. Oxford: Clarendon Press, 1939.

Santas, Gerasimos Xenophon. "An Argument against Explanations of Weakness." *Philosophical Review* (1966). Reprinted in Santas, Gerasimos Xenophon. *Socrates: Philosophy in Plato's Early Dialogues*. London: Routledge & Kegan Paul, 1979.

Scanlon, T. M. *What We Owe to Each Other*. Cambridge: The Belknap Press of Harvard University Press, 1998.

Schueler, G.F. *Desire: Its Role in Practical Reason and the Explanation of Action.* Cambridge: MIT Press, 1995.

Simpson, Evan. "Between Internalism and Externalism in Ethics." *The Philosophical Quarterly* 49 (1999): 201-14.

Smith, Adam. *The Theory of Moral Sentiments* (1759). Edited by D. D. Raphael and A. L. Macfie. Indianapolis: Liberty Fund, 1984.

Smith, Holly. "Culpable Ignorance." *The Philosophical Review*, XCII (1983): 543 -571.

Smith, Michael. "The Humean Theory of Motivation." *Mind* 96 (1987): 36-61.

—"Dispositional Theories of Value." *Proceedings of the Aristotelian Society*, Supplementary Volume (1989): 89-111.

—*The Moral Problem.* Malden, MA and Oxford: Blackwell Publishers, 1994.

—"Internal Reasons." *Philosophy and Phenomenological Research*, LV (1995): 109 -31.

—"The Argument for Internalism: Reply to Miller." *Analysis* 56.3 (1996): 175-184.

—"In Defense of *The Moral Problem*: A Reply to Brink, Copp, and Sayre-McCord." Ethics 108 (1997): 84-119.

Strawson's P.F. "Freedom and Resentment." *Proceedings of the British Academy* 48 (1962): 1-25.

Sturgeon's Nicholas. "What Difference Does it Make Whether Moral Realism is True?" *Moral Realism: Proceedings of the 1985 Spindel Conference. The Southern Journal of Philosophy* Supplement 24: 115-41.

Svavarsdóttir's Sigrún. "Moral Cognitivism and Motivation." *Philosophical Review* 108 (1999): 161-219.

Walsh, J. J. *Aristotle's Conception of Moral Weakness.* New York & London: Columbia University Press, 1960.

Watson, Gary. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." *Responsibility, Character, and the Emotions.* Edited by Ferdinand David Shoeman. Cambridge: Cambridge University Press, 1987.

Wiggins, David. "Weakness of Will, Commensurability, and the Objects of Deliberation and Desire." *Proceedings of the Aristotelian Society* (1979): 251-277.

—*Needs, Values, Truth.* Oxford: Basil Blackwell, 1987.

—"Moral Cognitivism, Moral Relativism, and Motivating Moral Beliefs." *Proceedings of the Aristotelian Society* (1991): 61-85.

Williams, Bernard."Persons, Character, and Morality." *Moral Luck*. Cambridge: Cambridge University Press, 1981.

—"Internal and External Reasons." *Moral Luck*. Cambridge: Cambridge University Press, 1981.

Wolf, Susan. "Asymmetrical Freedom." *The Journal of Philosophy*, LXXVII, No. 3 (1980): 151-66.

—"The Reason View." *Freedom Within Reason*. Oxford: Oxford University Press, 1990. Reprinted in *Agency and Responsibility: Essays on the Metaphysics of Freedom*. Edited by Laura Waddell Ekstrom. Boulder, CO: Westview Press, 2000.