August 2010

Kant and the Fact of Reason

Kenneth KH Chung
The University of Western Ontario

Supervisor
Dennis Klimchuk
The University of Western Ontario

Graduate Program in Philosophy

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

KANT AND THE FACT OF REASON

(Thesis format: Monograph)

by

Kenneth Chung

Graduate Program
in
Philosophy

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor	Examiners
Dr. Dennis Klimchuk	Dr. Anthony Skelton
Supervisory Committee	Dr. Corey Dyck
Dr. Samantha Brennan	Dr. Richard Vernon
Dr. Anthony Skelton	Dr. Sergio Tenenbaum

The thesis by

Kenneth Chung

entitled:

KANT AND THE FACT OF REASON

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date	Chair of the Thesis Examination Board

ABSTRACT

It is often thought that Kant abandoned his argument for the justification of morality in the *Groundwork of the Metaphysics of Morals* for a radically different argument in the *Critique of Practical Reason*. In the *Groundwork*, Kant appears to try to justify our commitment to the moral law on the basis of our freedom, but in the *Critique*, he tries to justify that commitment on the basis of what he calls the fact of reason. I assess and reject influential interpretations of both arguments as being philosophically unsound, and I propose, what I take to be, a novel and promising account of the fact of reason.

Keywords: Kantian ethics, Kant, moral philosophy, ethics, morality, the fact of reason, the motive of duty.

ACKNOWLEDGMENTS

I WOULD LIKE TO THANK the following persons:

My thesis supervisor, Dennis Klimchuk, who read many drafts of many versions of all my chapters. He always did so with care and patience. Meeting with him to discuss what I wrote was always something I looked forward to. That he is such a good philosopher didn't hurt either.

Professors Samantha Brennan and Anthony Skelton for being on my supervisory committee, for reading my dissertation, and for providing me with their helpful comments. I'm especially grateful to Anthony Skelton, whose comments and suggestions were extensive and useful.

Also, Professor Henrik Lagerlund for having read and provided comments on an early version of a chapter. Also to the members of my examining committee: Professors Richard Vernon, Corey Dyck, and Sergio Tenenbaum.

Greg Andres, for reading various chapters of my thesis at various stages, and for being so open to discussion of both the finer and rougher points of philosophy, sometimes related to my thesis and sometimes not, over the many years and many cups of coffee. My years as a PhD student would have been far less fun, far less enriching, and far less caffeinated otherwise.

Liz Sutherland, Julie Ponesse, and Amanda Porter who all read and provided comments on various parts of my thesis.

The grad students, faculty, and staff here at the Department of Philosophy. I have been lucky in the kindness and generosity I have been shown here.

Contents

LIST OF ABBREVIATIONS

A/B *Critique of Pure Reason*, by Immanuel Kant. English quotations are from the translation by Norman Kemp Smith, *Immanuel Kant's Critique of Pure Reason*. St. Martin's Press, 1965.

G *Groundwork of the Metaphysics of Morals,* by Immanuel Kant. English quotations are from the translation by Mary Gregor, *Groundwork of the Metaphysics of Morals.* Cambridge UP, 1998. References are to the page numbers of the "Akademie" edition.

CPrR *Critique of Practical Reason,* by Immanuel Kant. English quotations are from the translation by Mary Gregor, *Critique of Practical Reason.* Cambridge UP, 1997. References are to the page numbers of the "Akademie" edition.

T *A Treatise of Human Nature,* by David Hume. Oxford University Press, 1978, 2nd edition. Edited by L. A. Selby-Bigge and P. H. Nidditch.

EPM "An Enquiry Concerning the Principles of Morals." In *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, edited by L. A. Selby-Bigge, and P. H. Nidditch, Clarendon Press, 1975. 3rd edition.

Chapter 1

Kant, moral theory, and its justification

1.1 Introduction

Kant was situated at the end of the early modern era, and his work was a reaction to its two dominant philosophical movements: empiricism and rationalism. For Kant, Hume and Leibniz represented the dominant figures of their respective traditions. When trying to understand Kant's views, it is often useful to try to first understand his reasons for rejecting Hume's and Leibniz's views. This approach is useful not only with respect to trying to understand Kant's metaphysical and epistemological views, but also, I think, useful with respect to his ethical and metaethical views.

Today, debates in metaphysics, and in analytic philosophy more broadly, seem to assume a largely empiricist framework. Hume's descendants are also common in the landscape of ethics and metaethics. Not all of Hume's descendants would be happy being called sentimentalist, but they tend to share his naturalism. I believe that the problems posed by the empiricist and the rationalist traditions persist into much of contemporary philosophy, albeit mostly in a sharper, clearer language with a few changes in focus, but they are, I think, essentially the same.

Kant's rejection of both the rationalist and empiricist approach to ethics has not yet,

I think, been fully absorbed into contemporary philosophy. What we might learn from Kant's rejection of both Leibniz's and Hume's views will help us find a way of responding to their contemporary counterparts.

In the *Critique of Pure Reason*, Kant argued for, among other things, a sort of empiricism emphasizing its scope and its limits. He recognized that empiricism poses particular problems for the existence of morality: empiricist determinism threatens our freedom, empiricist ontology leaves little room for moral properties to dwell, and empiricist psychology explains that, despite our pretenses, we cannot be anything but selfish. Leibnizian rationalism also posed its own problems. Rationalists likewise held that the world was completely determined, but they also held peculiar views about the nature of moral truths and how we access them.

Morality, being concerned with the principles of right action and how they motivate us, appears to be puzzling if we accept a thoroughly empiricist or rationalist perspective. How can we be morally responsible for our actions if determinism is true and we could not have done otherwise? How can empirical, contingent facts about how things are tell us, by themselves, how things should be? And why should we care? In the *Critique of Practical Reason*, Kant answers that in practical contexts, these problems should not worry us, and that it is not only reasonable but demanded of us to accept the validity of the moral law as is. Once we know which action is the morally right one to choose, we are required to make that choice, despite philosophical worries about the nature of morality.

Kant articulated his moral theory in a intellectual culture dominated by the Leibnizian model via Christian Wolff. But Kant was also much influenced by the empiricists, in particular by Hutcheson and Hume. Both of these traditions, rationalism and empiricism, are echoed strongly in Kant's work. Their respective moral outlooks resonate in odd ways in Kant's own moral philosophy. Kant, however, was also influenced by Rousseau, who emphasized the capacity of ordinary people to act morally. Though Leibniz and Hume were

both egalitarians in their own respects, Rousseau's emphasis on the morality of ordinary people convinced Kant that the source of morality must lie deep within our equally shared human nature.

Kant tried to make good on this idea by incorporating it into a systematic moral philosophy that would avoid the mistakes he found in both rationalism and empiricism. The purpose of this introduction is to explain Kant's rejection of the rationalist and empiricist foundations of morality, thereby explaining his motivation to offer the kind of moral philosophy he did. And I hope that my project overall will help us better understand Kant's views about the justification of morality.

1.1.1 Kant's expectations

Kant had certain expectations of what a moral theory should do or be able to do.[1] I will try to explain his expectations, and they may strike some readers as plausible. They are, however, not uncontroversial. Many philosophers today reject them for a variety of reasons, so I don't intend to defend them. But by articulating Kant's expectations (or criteria), we might better understand why he rejected other moral theories and why he tried to locate morality in reason.

Like Rousseau, Kant admired the moral capacity of ordinary people—both their capacity to act from moral motives and their capacity to access genuine moral truths. For Kant, the basic truths of morality are not to be imparted to the masses by a select group of people, such as priests or theologians who might claim to have privileged access to morality. So the

[1] My use of the term "moral theory" runs over a contemporary distinction between normative and metaethical theory, or what J.L. Mackie called first order and second order moral views. Whether metaethics can be separated from normative ethics is a controversial issue. Ronald Dworkin (1996) thinks that they cannot be separated. Mackie (1990), for instance, thought that first and second order moral views were completely independent (16).

But Kant, like many early modern thinkers, does not adopt such a distinction. Rather, he tends to integrate metaethical concerns with normative concerns into a systematic whole. Stephen Darwall (1997) calls this kind of project "philosophical ethics."

first criterion is this: the correct account of morality must allow that basic moral knowledge is accessible to most everyone.

This leads to a second criterion: a correct moral theory must respect certain features that are implicit in what Kant calls "common cognition", which is just the term he uses in the *Groundwork* to refer to our ordinary moral thought (G 392). If moral knowledge and moral behaviour is within the province of what ordinary people think and do, our ordinary moral thinking must contain many elements of truth.[2] Kant thus proceeds "analytically from common cognition" (G 392), which means that he intends to derive the basic features of morality by analysing some of our basic moral concepts.

Kant's analysis of our ordinary moral thought leads him to two more criteria. The third criterion is this: that moral obligations are unconditionally binding. Kant writes, "Everyone must grant that a law, if it is to hold morally, that is, as a ground of an obligation, must carry with it absolute necessity" (G 389).[3] Our moral obligations are not optional, so to speak. Our moral considerations always trump other practical considerations. If it were in our self-interest to make a false promise but at the same time know that it is wrong to make a false promise, our self-interest ought to be defeated. Kant thought that this was a central conviction of our ordinary conception of morality.[4]

Here's the fourth criterion: moral considerations must be able to move us to act in

[2]I believe that most philosophers value our ordinary intuitions, but the extent to which philosophers value them varies greatly. Within moral philosophy, some contend that our ordinary moral thought often goes wrong, and that by relying on it, we might lead our lives with a certain moral complacency that is not, in truth, morally permissible. For instance, Peter Unger (1996), Peter Singer (1972), and Shelly Kagan (1989) all argue that we need to question our ordinary moral thought or behaviour.

[3]This way of phrasing it—using words such as "unconditional" and "absolute"—might be misunderstood to mean that certain actions are always wrong no matter the circumstances, and that there are never exceptions. Kant himself leads you to misunderstand him, because he once argued that we should never lie. He goes so far as to say that if a murderer came to your door looking for his intended victim, you are not morally permitted to lie to him ("On a Supposed Right to Lie from Altruistic Motives" 427). Kant's contention here is far-fetched, but it is distinct and independent from the criterion being considered here.

[4]It must be noted that many contemporary philosophers deny that moral considerations are always more important. Bernard Williams (1985), for instance, famously did not believe that moral reasons were always more important (184f). Susan Wolf (1982) also argued against moral perfection. And Michael Stocker (1976) argued that acting for moral reasons sometimes strikes us as wrong.

their accordance without the aid of extra motivation.[5] And for Kant, a correct moral theory must not only respect that condition but must recognize the value of so acting. Kant highly valued our capacity to act from the motive of duty, so much so that he seemed to think that an action has moral worth if and only if it was done from duty. Kant writes,

> [I]f an unfortunate man, strong of soul and more indignant about his fate than despondent or dejected, wishes for death and yet preserves his life without loving it, not from inclination or fear but from duty, then his maxim has moral worth. (G 398)[6]

We can see Kant's third criterion as motivating his fourth. Moral considerations would be rather useless and ineffectual if they were more important than other practical considerations but couldn't by themselves move us to act in their accordance. Suppose they were to conflict with all practical considerations, and that we recognized their superiority but at the same time were left unmoved to act in their accordance. Moral considerations would be a sham, unable to deliver on their supposed overriding authority. A moral theory must therefore allow for the possibility that a rational agent can be moved to act from moral considerations alone.

Now for the fifth criterion of an adequate moral theory: it ought to be able to give us moral guidance. In other words, it must help us adjudicate between conflicting claims, and it should be able to tell us whether an action is permissible or impermissible.

Kant also thought that we should never blindly accept any policy, and that we must always ask for its justification. But accepting a justification requires accepting further claims, and we must ask for the justifications of these further claims too. Suppose we

[5] As I've formulated it, this criterion is closely related to, but slightly different from, a widely accepted thesis known as "moral internalism", which is the view that moral judgments are closely connected with being motivated to act in accordance with them. But many philosophers—David Copp (1997), Shafer-Landau (2003), Svavarsdóttir (2006), for instance—reject moral internalism, often on the grounds that it is less consistent and less universal than its proponents think.

[6] Some have drawn from this the lesson that a person's action can have moral worth only if there are no other practical reasons to perform that action. But Kant intends no such implication. I don't wish to rehash the many reasons to reject such an interpretation. For an instructive account, see Herman (1993), ch. 1.

are told to follow a certain moral policy—say, that we should never steal—we should not simply follow this advice, but we must ask for its justification. If we are presented with a valid argument for why we should never steal, we must ask then for the justification of the premises in that argument too. An argument's validity, by itself, cannot justify our belief in its conclusion unless its premises are also reasonable.[7] And if we think, as Kant does, that there is *always* a principled way of adjudicating between different moral claims, then there can only be one supreme moral principle. So this is the sixth criterion—that there is only one such principle.[8] I should stress that this criterion is hardly unique to Kant. Many moral theorists also accept this; for instance, many utilitarians also accept it.[9]

Now, for the seventh criterion. Kant held that the basic moral principle is "not limited to human beings only but applies to all finite beings that have reason and will and even includes the infinite being as the supreme intelligence" (CPrR 5: 32). So it must hold not only for human beings, but also for God, angels, and intelligent extraterrestrials. In the second *Critique*, Kant arrives at this conclusion by examining what it means for a being to have a will and to be rational. But given the way I've set up the project, this criterion ought to be accepted prior to accepting that morality applies to all rational beings with a will. It is not clear to me what motivates this criterion, but Kant does seem to have independent reasons to hold this criterion. He writes,

> Even the Holy One of the Gospel must first be compared with our ideal of moral perfection before he is cognized as such; even he says of himself: why do you call me (whom you see) good? none is good (the archetype of the good) but God only (whom you do not see). But whence have we the concept

[7]Consider the following justification: Aristotle once stole from Jimi Hendrix, and since everything Aristotle did was wrong, it is therefore wrong to steal. We have a valid argument for the moral claim that it is wrong to steal, but we should hardly accept the claim for this reason.

[8]There is an important philosophical question of how we might justify such a principle, whether it might be self-evident, supported by non-moral claims, or supported in some other way. We might think that this is the key question about ethics, but it need not be answered here.

[9]There are, however, some philosophers, *moral particularists*, who reject this criterion. They doubt whether such a systematic method could capture the complexity and range of our moral practice. Jonathan Dancy (2004) is a notable defender.

> of God as the highest good? Solely from the idea of moral perfection that reason frames a priori. . . (G 408–409)

So in order to judge God good or any being as good, we would first need a notion of what it means to be good. Kant's point seems to be similar to Socrates' in Plato's *Euthyphro*: the concept of good (or right) is independent of God. But the relevant point here is that for Kant it is meaningful to consider God as *morally good*. On the other hand, for Kant, it would be meaningless to consider non-rational entities, such as rocks, plants, or most animals, as morally good. This provides some motivation for Kant to think moral principles apply to all rational beings, human or otherwise.

Another possible motivation comes from Kant's belief that morality must apply to any rational human being. And the word "any" here is important. Kant does not mean only that the moral law must apply to all existing rational human beings, but also that it must apply to all human beings that can ever possibly exist. It would not be enough for Kant if morality only applied to existing human beings, because it would otherwise always be possible that tomorrow a human being is born for whom morality is not a requirement. That would be an unacceptable consequence for Kant. One way of guaranteeing such a requirement is to require, as Kant does, that any intelligent being with a will must stand under the moral law.

And now for the eighth and last criterion: morality must not be an illusion. At the end of *Groundwork II*, Kant takes himself to have explicated what morality requires, but not to have shown that it is anything more than "a chimerical idea". He says that the first two chapters of the *Groundwork* have been "merely analytic" (G 445). What he says he needs to do is to prove that the Categorical Imperative—what he calls fundamental principle of morality as it applies to us human beings—is in fact true. And he attempts to prove that in the *Groundwork's* third chapter. Our first seven criteria could all be understood as a description of what morality must be like *if it existed*. In other words, even if we could be able to articulate a moral theory that respected the first seven criteria, there would be no

guarantee that morality is an actual force in our lives rather than merely a "phantom".

So in sum, here are the eight claims that Kant expects a satisfactory theory of morality to respect.

(1) Moral knowledge is accessible to most everyone.

(2) The correct account must respect the basic features of morality implicit in our ordinary moral thought.

(3) Moral obligations are unconditionally binding.

(4) Moral considerations must be able to move us to act in their accordance without the aid of extra motivation.

(5) The correct moral theory must be able to give us moral guidance.

(6) There is only one supreme moral principle.

(7) The supreme moral principle applies to all rational beings—human or otherwise.

And (8) Morality is not an illusion.

I started listing off these criteria as a means of qualifying the ambitions of my project in order to say that I cannot pretend to justify a moral theory above all others. But I still want to make a few more qualifications—this time on the criteria themselves. First, I do not pretend that these criteria are mutually exclusive. In fact, I have tried to show how some motivate other ones. Second, this is not an interpretation of how Kant saw his own project. He nowhere lists these criteria and then claims to seek a theory that satisfy these criteria. He may arguably be doing something similar in the first chapter of the *Groundwork*, where he lists three propositions, and then seeks to find the principle of morality that respects those propositions.[10] I have instead extracted these criteria from various passages. In some places, he is explicit about certain requirements. In others, he has criticized other views, and by trying to understand his criticisms, we can extract further criteria. But since he

[10]Kant doesn't explicitly mention the first proposition, but he mentions the second and third. The second is that "an action from duty has its moral worth *not in the purpose* to be attained by it but in the maxim in accordance with which it is decided upon", and the third is that "duty is the necessity of an action from respect for law" (G 399–400; emphasis in the original). The first presumably concerns the value of acting from the motive of duty.

nowhere makes the list of criteria as I have, my third qualification is this: I make no claim that these criteria are exhaustive. In fact, I know that they're not. For instance, I have not included the second and third propositions he lists in the *Groundwork*. But these two propositions are more controversial, and I believe that the ones I've listed are the central ones which motivate any Kantian moral theory. Third, I make no claim that the way I've listed the criteria is the only way to carve up the criteria. There could be better ways. Fourth, I only describe these criteria to help us better understand Kant's rejection of other views and his motive to give morality a basis in reason. I will not try to determine whether Kant own theory, in fact, succeeds or fails in meeting his own expectations.

There's another condition which guides Kant's whole moral project. I haven't included it in the eight criteria above, because this condition isn't motivated by Kant's considerations of moral theory alone. It is the condition that a satisfactory moral theory must be consistent with Kant's findings in the first *Critique*. While reading the *Groundwork* or the *Critique of Practical Reason*, we notice that many of his argumentative moves depend on claims that he takes himself to have proven in the *Critique of Pure Reason*. I won't attempt to defend the results of the first *Critique*, but I will keep that work in mind, for it sits in the background of Kant's moral theorizing.

1.2 Other moral theories

1.2.1 Leibniz's moral realism

Moral realism, as a thesis, is hard to define. Simply saying that it means that morality is real is woefully inadequate because there is no consensus for what is to count as real. Let me then define two versions of moral realism: a *minimal* moral realism and a *robust* moral realism. Minimal moral realism is the view that moral judgments express beliefs and that

some moral beliefs are true. Kant is undoubtedly a minimal moral realist. Minimal moral realism barely qualifies as a metaphysical view—the first part is a linguistic thesis, which is called *cognitivism*. It simply states that our expression of a moral judgment is, in general, the kind of thing that can be true or false.

The second part asserts only that some moral judgments—in particular, judgments such as "*X* is good" or that "one ought to *A* in circumstances *C*"—are true. These claims, neither separately nor together, entail any substantive metaphysical view, because they do not demand any particular notion of truth. Moreover, some theories of truth are neutral about metaphysics; for instance, there are deflationary theories of truth, which hold that "for a sentence to possess truth-conditions it is sufficient that it be disciplined by norms of correct usage and that it possess the syntax distinctive of declarative sentences" (Miller 2007, 197). One could thus accept minimal moral realism, a deflationary theory of truth, and still not make any metaphysical claim. What I call *robust moral realism* is a more clearly metaphysical view, and the one that is more generally associated with the term "moral realism" simpliciter. Russ Shafer-Landau is a current exponent of this type of moral realism, and he describes it as

> the view that says that most moral judgments are beliefs, some of which are true, and, when true, are so by virtue of correctly representing the existence of truth-makers for their respective contents. Further, and crucially, true moral judgments are made true in some way other than by virtue of the attitudes taken towards their content by any actual or idealized human agent. (210)

One robust moral realism from the early modern period is that of Leibniz's. We can get a glimpse of Leibniz's take on moral truths by looking at his rejection of divine voluntarism. *Divine voluntarism* is the thesis that there are moral truths but that they are dependent on God's will. Divine voluntarists generally hold that God, by an act of will, makes the moral truths. And though Leibniz shares with the divine voluntarists a belief in God, he outright rejected voluntarism for two reasons.

Very early on in the *Discourse in Metaphysics*, in §2, Leibniz begins his attack on the thesis of divine voluntarism. The thesis of divine voluntarism, Leibniz says, would imply that God's decisions regarding what moral laws to make cannot be guided by any moral principle, since there are no moral principles before God declares them to be so. This result is unacceptable for Leibniz, and he believes that it would be contrary to any reasonable understanding of God. Leibniz has at least two reasons why this is unacceptable.

First, divine voluntarism would mean that we could not love God. Any such love would be irrational. Thinking of God as the maker of morality places him outside our moral community, making it senseless to love him. Second, it implies God could not be good, let alone perfect. We may not know why God makes the moral laws as they are, but we do know that it cannot be out of concern for good and evil, for such notions are moral notions and would not yet exist. According to Leibniz, this would make God indifferent to good and evil, and such indifference can hardly be praised. If God could have made the world differently, then any such world would also be just as good, simply because God so decrees it. Leibniz writes,

> It is a happy necessity which obliges wisdom to do good, whereas indifference with regard to good and evil would indicate a lack of goodness or wisdom. (*Theodicy* §175)
>
> They [the divine voluntarists] deprive God of the designation *good*: for what cause could one have to praise him for what he does, if in doing something quite different he would have done equally well. (*Theodicy*, §176)

This makes our praise and love for God peculiar and perhaps incoherent.

What this means about the moral laws is that they exist outside God, and thus, also outside the actual universe as we know it. In Leibniz's view, the actual universe was made actual by God upon His choosing it from a set of possible worlds. God's decision to choose this universe to become actual is influenced in part by the moral laws. In order to praise God for his perfection, he must have chosen the most perfect universe to be made actual.

Perfection, for Leibniz, is achieved by reaching a balance of various eternal principles, some of which are moral principles. But we human beings can never know *how* the universe we live in is perfect, because all the principles relevant to God's decision of making this universe actual are not available to us, but only to God. We do however know *that* it is perfect. We can come to the conclusion that it is perfect based on premises that do not explain why this particular world was chosen. Our knowledge that the world is perfect is separate from the issue of how it is perfect. This is not unusual. If we have good reason to trust A to select the best X from a set of X's, then we can know that after A has so chosen, that it is the best X, but we may not know why it is the best X, because we may not know what principles guided A's decision. For instance, I know that the American Olympic committee will choose an excellent pitcher for their Olympic baseball team. I know very little about baseball and about the characteristics of baseball pitching excellence. I expect the reliability, speed and accuracy of a pitcher's throw to be relevant qualities of an excellent pitcher, but I don't know all of them. But I'm nonetheless confident that the American Olympic team will have an excellent pitcher. Thus I don't know *why* the pitcher will be excellent, but I know *that* he will be. Whenever we rely on experts to choose the best X, we know that it will be very good, but we may not know why it is very good. The analogy isn't perfect, because God will choose the *best* possible world, not just an excellent one. I said that the American Olympic team will have an excellent pitcher, instead of the best one, because human beings are imperfect and so I do not know whether the best pitcher will be chosen. But since I presumably know that God is perfect, his decision cannot suffer from human errors.

According to Leibniz, for us human beings, our knowledge of right and wrong supposedly depends on our perceptive ability. Leibniz believed, as was common in the early modern era, that there was one faculty, namely perception, under which falls all of what we would call beliefs, sense perceptions, feelings, intuitions, suspicions, etc. The differences

between sense perceptions, feelings, intuitions, and so on depend on how a perception is manifested, and not on which faculty they issue from. So our perception of right and wrong is simply a different manifestation of our one perceptual sense, just as our visual perception is. All this to say that what is morally right, for Leibniz, is something that is out there to be perceived, and it is not simply a function of the attitudes we human beings take. Furthermore, when our moral judgments are correct, it will be because we have perceived accurately. Leibniz is clearly a robust moral realist.

So why does Kant reject Leibniz's moral framework? Kant's most explicit reason depends on his rejection of Leibniz's account of freedom. Leibniz endorsed compatibilism, which is the thesis that there is no inconsistency between freedom and determinism. But he endorses a peculiar kind of compatibilism, due in part to his peculiar reasons for accepting determinism and free will.

Empiricist versions of determinism tend to get their impetus from a plausible claim about causality: any event is sufficiently determined by the past and the laws of nature. (Other kinds of determinism may also depend on a claim of causality. For instance, a dualist might be a determinist if he accepts a slightly different causal claim, one which says that any event is determined by the past, the laws of nature, and the laws of the mind.) But Leibniz's belief in determinism arises less from a belief about causation than about truth. Suppose we accept the theory that every proposition is either true or false. This might seem uncontroversial until we consider statements about the future. Consider the sentence: "it will rain here tomorrow." The theory in question implies that the proposition expressed by this sentence is either true or false; there is no third truth value or indeterminate truth value. We may not know what its truth value is until tomorrow, but it already has one. If it is true, then it is a fact about tomorrow which makes it so. And if it is false, it is a fact about tomorrow which makes that so. Either way, its truth value is already determined today, even if we don't know what it is. And this is the thesis of determinism: any claim about the

future is already true or false, even if we don't know which.

Determinism seems inevitable if you subscribe to Leibniz's view of truth. He held that truth was a matter of concept-containment; a proposition is true if and only if its subject contains its predicate. In his essay "On Freedom", Leibniz wrote,

> . . . it is common to every true affirmative proposition, universal or particular, necessary or contingent, that the predicate is in the subject, that is, that the notion of the predicate is involved somehow in the notion of the subject. (95)

This means that, in modern parlance, that all truths are analytic. Since this applies to any true sentence, it must also apply to claims about the future. The sentence "John will eat candy tomorrow" has a determinate truth value today. And if it is true, then the subject, *John*, contains the predicate, *will eat candy tomorrow*. So the subject already contains the predicate, even though the predicate describes an event that has yet to happen. And since every event that occurs or will occur in the universe can be described by a true—and thus analytic—claim before it even happens, determinism must be true.

The trickiest part for any compatibilist theory is to explain free will in a way that doesn't require the falsity of determinism. For Leibniz, freedom requires only three conditions: spontaneity, intelligence, and contingency. Leibniz writes,

> I have shown that freedom. . . consists in intelligence, which involves a clear knowledge of the object of deliberation, in spontaneity, whereby we determine, and in contingency, that is, in the exclusion of logical or metaphysical necessity. (*Theodicy*, §288)

For Leibniz, both God and human beings are free, and in the same sense; the difference is only a matter of degree.

To say that a truth is contingent means that it isn't necessary. And according to Leibniz, a truth is necessary if and only if its negation produces a contradiction (*Discourse on Metaphysics*, §13). So there appears to be a tension in his view: he believes that all truths

are analytic, but he believes that some truths are contingent. We modern readers may balk because we often think that what is analytically true is also necessarily true. Leibniz denies this; he believes that some analytic truths are contingent. But it's hard to see how this can be. Suppose that "John will eat candy tomorrow" is true. Then by his own theory of truth, this means that the concept of *John*, contains the predicate, *will eat candy tomorrow*. But this proposition is presumably contingent, which would mean that it's not contradictory to deny it. But if the concept of *John* contains the concept of *will eat candy tomorrow*, it seems to follow that it would be self-contradictory to assert that it is possible that John not eat candy tomorrow; in other words, it seems that the sentence "John will eat candy tomorrow" is necessarily true. Since contingency is important to Leibniz, he must find some way of explaining how analytic truths can be contingent.

Leibniz seems to employ a few different strategies to deal with this problem. He sometimes appeals to the distinction between hypothetical necessity and absolute necessity, allowing him to say that it is contingent that John will eat candy tomorrow because it is contingent that John exists (*Philosophical Essays*, 69–70). There is no absolute necessity in either of these claims. It is only hypothetically necessary that John eat candy tomorrow—necessary on the hypothesis that John exists. At other times, Leibniz seems to think that there are contingent connections within a concept; so that it would be contingent whether the predicate *will eat candy tomorrow* can be found in the concept *John.* I'm not sure how fruitful either of these strategies is, and according to Nicholas Jolley, "the problem of interpreting Leibniz's attempts at accommodating contingency in his philosophy is one of the most difficult and controversial in Leibniz scholarship" (135). Suffice it to say that it is far beyond the scope of this work to explore this issue in more detail.

Now to say that an action is free also requires that it be performed by an intelligence, which means a being with an intellect. For Leibniz, the intellect is the faculty of perception, where perception is understood to be much broader than simply sense perception. We

have perceptions of what is good, perceptions of what is beneficial to us, and so on. And for Leibniz, our perceptions can be clear and distinct, confused and obscure, and so on. Leibniz writes,

> Our knowledge is of two kinds, distinct or confused. Distinct knowledge, or *intelligence*, occurs in the actual use of reason; but the senses supply us with confused thoughts. (*Theodicy* §289)

The action we choose to take is always the one which we perceive to be most good. Leibniz writes,

> For we can only will what we think good, and the more developed the faculty of understanding is the better are the choices of the will. And, in the other direction, in so far as a man wills *vigorously*, he determines his thoughts by his own choice instead of being determined and swept along by involuntary perceptions. (*New Essays* 180)

And for Leibniz, only God and human beings are capable of such perceptions. Moreover, our action is free to the extent that the action we choose is, in fact, the most good.

It follows that for Leibniz, freedom is a matter of degree. Since intelligence is a matter of degree, so too is freedom. Some of us, on average, have a better perception of the good than others do. God then is perfectly free because his perception is flawless. He writes,

> . . . the question is not about whether his legs are free or whether he has room to move about, but whether he has a free mind and what that consists in. On this way of looking at things, intelligences differ in how free they are, and the supreme Intelligence will possess a perfect freedom of which created beings are not capable. (*New Essays* 181)

Our acts of *deliberation* are simply our attempts to see clearly. And we are virtuous to the extent that we see clearly. Very few philosophers today, if any, accept Leibniz's compatibilist view. If we can manage to ignore the difficult problem of reconciling the contingency of events with the analyticity of true statements that describe those events, we still have to manage to understand how a person's action can be spontaneous yet determined.

Kant famously calls the kind of compatibilism Leibniz defends "a wretched subterfuge" (CPrR 5: 96). Kant writes,

> . . . all necessity of events in time in accordance with the natural law of causality can be called the *mechanism* of nature, although it is not by this that the things which are subject to it must be really material *machines*. Here one looks only to the necessity of the connection of events in a time series as it develops in accordance with natural law, whether the subject in which this development takes place is called *automaton materiale*, when the machinery is driven by matter, or with Leibniz *spirituale*, when it is driven by representations; and if the freedom of our will were nothing than the latter (say, psychological and comparative but not also transcendental, i.e., absolute), then it would at bottom be nothing better than the freedom of a turnspit, which, when once it is wound up, also accomplishes its movements of itself. (CPrR 5: 97)

Here Kant puts together Leibniz's account of compatibilism with empiricist versions of compatibilism, and he rejects both on the same grounds. Both are versions of what is sometimes called classical compatibilism, which holds that all we need to be free is that we have the power or ability to do what we desire and that we suffer from no constraints or impediments (Kane 13). Kant mocks this kind of freedom.[11] He says that such theories of freedom use the word "free" in the same way we use it when we describe "that which a projectile accomplishes when it is in free motion"; free, because there are no forces acting on it other than gravity (CPrR 5: 96). He calls the classical compatibilist conception of freedom "comparative freedom', because on this view we only understand an action as free by comparing it with constrained actions. But for Kant, if a human being is *caused* to perform his actions by his desires, then the series of events from desire to action have a mechanical necessity. According to Kant, a human being on Leibniz's view cannot amount to anything more than a wind up device.

But the heart of Kant's rejection of classical compatibilism is that it would make moral

[11] It's unclear the extent to which Kant addresses Leibniz's peculiar conception of freedom as involving an intellectual perception of what is good. Nonetheless Leibniz seeks to find a kind of freedom within a determinist framework, and for Kant, this will never do.

praise and moral blame impossible, or at the very least, unreasonable. It would be unreasonable to blame someone for an action if we also regard him as not having been able to have done otherwise. Kant is appealing to a version of what has sometimes been called the *Principle of Alternative Possibilities*:

> Persons are morally responsible for what they have done only if they could have done otherwise. (Kane 80)

If determinism is true, then any action a person performs must have happened. No one could have performed an action different from the one performed. Kant considers the case of a man who commits an act of theft, but is caused to do so by natural necessity and so could not have done otherwise. He asks, "how, then, can appraisal in accordance with the moral law make any change in it and suppose that it could have been omitted because the law says that it ought to have been omitted?" (CPrR 5: 95) Kant's question is this: how can recognizing an action as right or wrong affect our actions if we were already determined to act in particular way? That is, what difference does morality make in our behaviour? For Kant, morality requires that we conceive of ourselves and others as having the freedom to choose between various courses of action. In particular, it requires conceiving of ourselves as being able to have performed actions other than the ones we did. Kant seems to think that moral assessment of our actions is pointless if we could not have done otherwise.[12] For Kant, Leibniz's compatibilism, and classical compatibilism more generally, makes morality an illusion.

But Kant's rejection of Leibniz's moral view consists in much more than a rejection of his view of freedom. Though he doesn't explicitly mention it in the *Groundwork* or

[12]This line of thought depends on the principle of alternative possibilities (PAP), which is highly controversial in free will debates. Robert Kane (1996) defends it, but Harry Frankfurt (1969) famously argued against it. Nomy Arpaly has even argued that PAP is not as central to our ordinary conception of moral responsibility as we are led to think, because we often take pride in not being able to have done otherwise (2006, chap. 2). For instance, we can think of Daniel Dennett's example of Martin Luther who accepts moral responsibility for his action when he says "Here I stand. I can do no other" (2003, 117).

in the second *Critique*, Kant would reject much of the metaphysics and epistemology that underlies Leibniz's view. For instance, Kant rejects the possibility that our perception is able to give us any information unless it is conditioned by space and time. That is, all perception, for Kant, is empirical perception. We do not 'perceive' right and wrong, and our perceptions do not move us to act. Kant would also reject the framework of a set of possible worlds, where God chooses the best of such worlds to make actual. And he would reject the principle of sufficient reason as an organizing principle of the way the universe works. Kant would likely think that Leibniz is illegitimately applying principles to objects considered outside of space and time, when they can only be used for objects considered within them. That is, Leibniz is trying to apply metaphysical principles to things in themselves, in one case to God and in the other to the grounds of the existing universe.

But let us see if Leibniz's moral theory fails to address the criteria that I've set out at the beginning of this chapter. There is a glaring omission in my account of Leibniz's view. I have yet to make any mention of Leibniz's fundamental moral rule. Leibniz does not seem to be explicit about what it is. However, he writes,

> Finally, under this perfect government no good action would be unrewarded and no bad one unpunished, and all should issue in the well-being of the good, that is to say, of those who are not malcontents in this great state, but who trust in Providence, after having done their duty, and who love and imitate, as is meet, the Author of all good, finding pleasure in the contemplation of His perfections, as is the way of genuine 'pure love,' which takes pleasure in the happiness of the beloved. (*Monadology*, §90)

Leibniz implies that our duty is to love and imitate God. This leads Jolley to conclude, given that God "manifests his goodness in creation by seeking to maximize the happiness of minds", we also ought to maximize the happiness of minds (183). Jolley writes, "Leibniz is thus close to the modern doctrine of utilitarianism which holds that the maximization of happiness is the fundamental rule of morality" (183). So Leibniz does have a fundamental moral principle, and it resembles the principle of utility. Most of my discussion of Leibniz

has not focused directly on his moral theory, but on how his metaphysical view leads him to adopt his peculiar moral view. This is mostly a result of Leibniz's own theorizing. He is not traditionally regarded as a moral philosopher, but he frequently ends his works with considerations in ethics even if they begin with considerations in metaphysics. The *Monadology* is one such example (Jolley 176ff). We would have to do some work to articulate a full moral theory, but I doubt that we could ever get a free standing moral theory, by which I mean one that does not depend on his metaphysical views. And since Kant rejects even the basic methodology of Leibniz's metaphysics, there is little reason for Kant to accept his ethics.

But there is a peculiar difficulty for moral motivation on Leibniz's view. On Leibniz's view we are *always* moved to act in accordance with that which we conceive of as good. More specifically, Leibniz thinks that we can only be motivated by the prospect of pleasure or pain. Pleasure and pain are caused by our perceptions of perfection, goodness, and harmony. Because we take pleasure in perceiving goodness, we are motivated to bring it about. When we act wrongly, for Leibniz it can only be the result of inaccurate perception of what is pleasurable. The French jurist Jean Barbeyrac raised the following objection against Leibniz's account of moral motivation. Basically, Barbeyrac says that Leibniz's account of morality misses a fundamental aspect of what it means to be moral. According to Barbeyrac, there is an important distinction between being honest and being useful. For Barbeyrac, we sometimes do the morally correct action simply because we think it is right, and not because of any reward. We may not expect any pleasure from doing what is right (Schneewind 256-257). Leibniz's account does not allow for this possibility.

This objection slightly begs the question against Leibniz, because it assumes that there is a distinction between moral obligation and self-interest, which is a distinction that Leibniz deliberately blurs. But I think Barbeyrac still has a point: it is part of our ordinary common moral thinking that moral duty doesn't always align with what is in our self-interest.

Leibniz's view doesn't respect that fact, and it's an example of how Leibniz's moral theory conflicts with our ordinary common moral ideas, which conflicts with Kant's second criterion. Recall also that the fourth criterion is that moral considerations alone must be able to move us to act without the aid of extra motivation. Leibniz's theory satisfies this criterion only by claiming that moral considerations are inherently self-interested and that we are only ever motivated by self-interest. But this cannot place the same value, as Kant does, on acting from the motive of duty. Leibniz's moral theory turns out to satisfy the fourth criterion only in letter, but not in spirit.

1.2.2 Hume's moral theory

David Hume held that our moral convictions are based ultimately on our feelings. He offered at least two arguments for this conclusion. The first argument can be found in this famous passage of his:

> Take any action allow'd to be vicious: Wilful murder, for instance. Examine it in all lights, and see if you can find that matter of fact, or real existence, which you call *vice*. In which-ever way you take it, you find only certain passions, motives, volitions and thoughts. There is no other matter of fact in the case. The vice entirely escapes you, as long as you consider the object. You never can find it, till you turn your reflexion into your own breast, and find a sentiment of disapprobation, which arises in you, towards this action. Here is a matter of fact; but 'tis the object of feeling, not of reason. It lies in yourself, not in the object. So that when you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it. (T 468–469)

Hume's point seems to be that we cannot find the wrongness of an action in the action itself. We can only find it in our response to the action—in a feeling of disapproval when we consider the act. Thus, the rightness or wrongness of the action is not in the action itself but in our feelings.

Hume's second argument begins with the premise that reason can only tell us what is true or false, and that reason can never move us to action or even move us to praise or condemn an action (T 415, 458). Reason, for Hume, only concerns relations of ideas or matters of fact. And neither being aware of a relation of ideas nor being aware of a matter of fact can, by itself, ever move us to action. Hume believes instead that only desires can move us to action. He writes that "it can never in the least concern us to know, that such objects are causes, and such others effects, if both the causes and effects be indifferent to us" (T 414). In other words, we must already have a desire or an aversion to an object (or its causes or its effects) for knowledge of its causes or effects to motivate us. Reason alone, for Hume, "can never produce any action" (T 414). But Hume takes it as plain that our moral convictions do move us to action, and so our moral convictions must, at bottom, be a kind of desire. "Morals", according to Hume, cannot then "be deriv'd from reason" (T 457).

Wrapped up in this argument against reason as the basis of morality is a partial defense that morality must be found instead in our feelings. But Hume doesn't have the simplistic view that X is good if and only if we like X. Rather, Hume has the view that X is a virtue if and only if we as spectators, taking up "steady and general points of view", feel approval towards X (T 581–2). And likewise, X is a vice if and only if we as spectators feel disapproval towards X. (It should be noted that Hume takes morality to be first about character traits, in particular about our virtues and vices, and our moral evaluation of people's actions derives from our judgments of their characters (T 477).) Taking up the general point of view when considering a person's trait requires considering him from the perspective of an impartial observer, as if the person's actions neither benefits nor harms us. And if we have a "pleasing sentiment of approbation" upon considering a person's mental quality, we call that quality "virtue"; we call it "vice" if we have a displeasing sentiment of disapproval (EPM 289).

For Hume, this capacity for approval or disapproval is not a product of reason, but rather the product of a moral sensibility. It is unclear to me whether Hume regards our moral sensibility as being distinct from our other sensibilities, or whether he regards it as derivative of another sensibility, namely our capacity for sympathy. As I understand it, Hume believes that our feeling of approval or disapproval is a kind of sympathy that we have towards the person and the people around him. But our sympathies can be biased, say by favouring fellow nationals, unless we regulate them by adopting the general point of view (Brown 25). It thus seems as if sympathy is the prime impetus of our moral judgments. But at the same time, it also seems to me that we need some additional impetus to adopt the general point of view in the first place, without which we only have unregulated sympathy. And maybe this additional impetus might arise from a distinctly moral sensibility. I don't really know, but either way, Hume thinks that we have a motivating sensibility that does not depend on the prospect of a person's actions benefiting or harming us.

Hume's moral theory has the resources to satisfy most of the criteria I've set out at the beginning.[13] But it doesn't satisfy all of them. Most obviously, Hume's theory denies that morality applies to all rational beings. As David Fate Norton puts it, for Hume, "as far as any of us knows or can know, morality has to do only with human beings and human affairs. We do not know what is expected of higher beings; our reason cannot reach such heights" (156). For Hume, we cannot say that our moral principles apply to all rational beings, human or otherwise, because we have no idea what other rational beings are like. We do not know and cannot know whether other rational beings will have a similar moral sensibility—we do not know whether they will take up the general point of view, whether

[13]It is compatible with claiming (1) that moral knowledge is accessible to most human beings, (2) that moral truths reflect features implicit in our ordinary moral thought, and (4) that moral considerations alone can move us to act. Hume's moral theory, insofar as it depends on this mechanism of an impartial spectator, has the resources to meet (5), give us moral guidance; and that might be enough to meet (6) as well, that there is one ultimate procedure for deciding between right and wrong. It's less clear whether it can show (3), (7) or (8).

they experience a similar sentiment of approval at the same set of character traits. That all (or most) human beings have the same moral sensibility is, for Hume, a fact, which we only learn from experience.

But it isn't clear to me that this criterion, the seventh, is reasonable—do moral principles apply to any possible rational being, human or otherwise? The original reason for thinking they do was that otherwise it would be senseless or nonsense to say that God is good. But it seems to me that it is one thing to judge whether someone's character or intention is good or bad, and another thing entirely for that person to consider whether his own character or intention is good or bad. In other words, Hume's view still allows us to say that God is good, provided we have a sufficiently pleasant sentiment of approval contemplating Him and His actions from the general point of view. And this does not require thinking that God must obey the same moral code as we do. It is no doubt odd to condemn a person's actions if we do not think that he was obligated to act otherwise. After all, we tend not to judge natural disasters as moral failures on the part of the universe. But as odd as it is, I'm not sure it's a contradiction. Either way, I am uncertain about this criterion.

The most serious problem for Hume's theory is that it fails to satisfy the third criterion, which states that a moral theory must explain, or at least be consistent with, the requirement that moral obligations are unconditionally binding. Let us suppose that we judge an action to be wrong. On Hume's view, this means that we have a certain feeling towards that action, albeit one we have only when we adopt a general point of view. But it is possible that we can, at the same time, have a conflicting desire. Perhaps performing that wrong action will help us satisfy a desire. So we can have a conflict—one between, say, an aversion, which arises from a feeling of disapprobation towards a particular action A and a desire for A's potential benefits. Can Hume's theory respect the idea that the first feeling outweighs or outranks the second? More generally, can it respect the idea that moral considerations override all other practical considerations? We cannot simply say that our moral desire

is always stronger than other conflicting desires, because otherwise people would always perform for moral reasons—and that's simply not true. We cannot say that our moral desire is generally stronger than other desires, because this still does not mean that moral considerations outweigh or outrank our other desires. We might try to adopt a self-reinforcing solution and suggest that from the general point of view, we always approve of those acting in accordance with the general point of view. But this appeal smacks more of circularity than of recursion, for we could never get to the conclusion that the general point of view is more trustworthy without already accepting it.

Hume famously thought that we can never find universal or necessary truths from an investigation of nature. All of our experiences only add up to a finite lot, and according to Hume, it would be invalid for us to infer any universal or necessary claim from a finite set of experiences. But moral theory, for Hume, is the result of an investigation of nature; in particular, it is an investigation of human nature. So moral theory should not be able to yield any genuine universal or necessary claim, and Hume does come to this conclusion. But moral theorizing usually involves certain modal claims: we *must* do this, it is *obligatory* that this is so, it is *impermissible* for anyone to do such and such, and so on.[14] The third criterion I attributed to Kant also involves a similar kind of modal claim: that moral obligations are *unconditionally binding*, which means that moral considerations *always trump* other practical considerations. But from an investigation of nature, we cannot hope to find any such modality, or any such unconditional importance. Hume concludes from all this that we don't have any such modality. We don't have genuinely necessary moral claims. Claims that begin with phrases such as "we must" or "it is obligatory that" can never, strictly speaking, be true. We can however get feelings of obligation and feelings of necessity, and those, for Hume, should suffice.

[14]These express claims of *deontic* modality, rather than metaphysical or epistemic modality.

Like Hume, Kant thought that we can never get any necessity from an investigation of nature. But unlike Hume, Kant didn't think that feelings of obligation would suffice. If we regard our obligations as feelings then we cannot, at the same time, regard them as unconditionally binding. Understood as feelings, they are too contingent and too dependent on empirical facts about human beings to be either necessary or universal. So both Hume and Kant agree to a large extent about what we can find in a study of human nature; they both agree that neither necessity or universality can be found by empirical means. But whereas Hume concludes that moral theory is not concerned with necessary or universal truths, Kant concludes that moral theory must require more than what can be found empirically.

Writing about a slightly different theory, namely, a theory that tries to base our moral obligations on our own self-interest, Kant writes,

> Suppose that finite rational beings were thoroughly agreed with respect to what they had to take as objects of their feelings of pleasure and pain and even with respect to the means they must use to obtain the first and avoid the other; even then they could by no means pass off the *principle of self-love* as *a practical law*; for, this unanimity would still only be contingent. The determining ground would still be only subjectively valid and merely empirical and would not have that necessity which is thought in every law, namely objective necessity from a priori grounds… (CPrR 5: 27)

Kant is objecting to a moral theory based on the principle of self-love on the grounds that it could at best only achieve unanimity, and that it could not produce any necessity or universality. Hume's attempt to articulate a moral theory entirely on human nature faces the same difficulty. It can at best only achieve a unanimous agreement on what is good and what is bad, but it could never give us a practical law, according to Kant, because it can never give us universality or necessity. Moral theory thus cannot solely be about human nature; we must look elsewhere for the universality and necessity of moral claims.

1.3 Kant's project

Kant tries to meet his criteria by locating morality in reason. Reason, after all, is something all human beings have, and that fact could explain why moral knowledge is accessible to everyone. A moral theory based on reason, because everyone has it, also has the resources to respect the basic features that we think morality has. Also, for Kant, we can never get universality (or necessity or normativity) from experience, so reason becomes an obvious place to look for the universality and normativity of morality.

Kant's criterion that moral considerations must be able to move us to act in their accordance without the aid of extra motivation suggests that what makes something obligatory (or right) and what motivates us to act morally should be one and the same thing. Hume and many Humeans also accept this result, but they take this as grounds for thinking that what makes something obligatory must be a desire or attitude. But I believe that they come to this conclusion, partly because they are already looking for a certain kind of explanation—a deterministic explanation of human behaviour

There are two ways of explaining our actions. Here's the first way. We can look for a purely scientific explanation of a person's action where our goal is to describe the past and the laws of nature which would sufficiently determine that person's action. For instance, we may refer to the person's beliefs and desires, to the circumstances of the moment, and appeal to the folk psychological laws, to explain why he did what he did. Or we may refer to the neurophysiological states of the person's body and appeal to the laws of neurophysiology in order to explain why he did what he did. If a person does something different from what we expect, we would look for a belief, or a desire, or a neurophysiological element that may have escaped our notice, or perhaps we may revise and refine our folk psychological laws or our neurophysiological laws. That is, we assume that the past and that the laws of nature sufficiently determine every event that ever happens,

including any person's action, and use that assumption as a methodological principle to guide our inquiry for behaviour explanation. To put it another way, we assume the truth of determinism to help guide us in our search for good explanations. I believe that this assumption motivates the Humean's inquiry. I'm not sure that there is anything inherently wrong with this. I think Kant accepts this assumption too, but he would confine its use to certain domains of inquiry.

Kant, as I mentioned above, does not think that determinism is compatible with morality; in particular, he thinks it would make our appraisals of blame and merit unreasonable. We do want to praise and blame people, but we cannot do so without giving at least something resembling an explanation for their actions. We need to know why a person did what he did in order to think that he is praise or blame worthy. For Kant, we cannot assign blame or praise to a random action. We ought instead to adopt the second way of thinking about and explaining a person's actions: we look at a person's *reasons* for acting. Kant thinks that we must consider a person's reasons for actions in order to be able to hold him morally responsible for his actions. For instance, acting for selfish reasons might merit blame. So I believe that while Kant holds with the Humeans that what makes something obligatory should be the same thing that motivates us to act morally, he differs on what that is. For Kant, it is a kind of reason that both makes something right and moves us to act morally.

Because Kant believes that the moral obligations must be unconditionally binding, a moral reason for acting cannot depend on any empirical condition for its normative force. And so Kant concludes that the moral law must arise from reason itself, and not from any contingent feature of humanity. He writes,

> . . . when I think of a *categorical* imperative I know at once what it contains. For, since the imperative contains, beyond the law, only the necessity that the maxim be in conformity with this law, while the law contains no condition to which it would be limited, nothing is left with which the maxim of action is to conform by the universality of a law as such. (G 420–1)

Kant seems to think that if we deprive all our obligations of anything conditional, the only obligation left is the idea of a universal law in itself. That is, we try to think of what we are obligated to do that does not depend on any interest we happen to have, and what we are left with is the mere form of a law. Kant thus introduces his fundamental principle of morality, which he calls the Categorical Imperative: "act only in accordance with that maxim through which you can at the same time will that it become a universal law" (G 421). The right reason for action, then, is one that passes the test of the Categorical Imperative. We consider an action and our reason for performing it, and then we ask whether we can will that reason as a universal law and at the same time choose to act on it. This is meant as a test of our reasons for acting: if our reason for performing a particular action does not pass this test, then we are not permitted to act on that reason; if it does, we are so permitted. And if our reason to refrain from performing a particular action fails to pass the test, then we ought to perform it. As such, the Categorical Imperative can give us moral guidance.

We can now see how Kant's moral theory might go some way to meeting his expectations: by basing morality in our reason, he seems to have the resources to explain (1) how moral knowledge might be accessible to most everyone, (2) why the basic features of morality, as we ordinarily conceive of them, might be correct, (3) how moral obligations might be unconditionally binding, (4) how moral considerations alone might move us to act, (5) how moral principles might gives us guidance, (6) why there is only one supreme moral principle, which he calls the Categorical Imperative, and (7) why moral requirements, being products of reason, apply to any possible rational being, human or otherwise. Whether his moral theory does, in fact, have the resources to meet his expectations is not an issue I wish to settle. Rather my point is that we can see why, given his expectations, Kant tries to locate morality in reason.

Kant's last task is to satisfy the eighth criterion; that is, to show that morality is not an illusion. In particular, he needs to show that the moral law does in fact apply to us.

This means several things: that we can in fact choose to act in accordance with it, that we can in fact act from moral considerations alone, and that the moral law, in the form of the Categorical Imperative, does indeed bind us.

In the *Groundwork*, Kant attempts to do all this in the third chapter. He begins with the assumption that we are rational agents. From that assumption, he argues to our being negatively free, then to our autonomy, and finally to our subjection to the Categorical Imperative. And if this argument works, Kant shows that insofar as we take ourselves to be rational agents, we stand under the Categorical Imperative. In my second chapter, I attempt to offer what I think is the best interpretation of Kant's argument in the *Groundwork*, but I argue that it fails.

In the *Critique of Practical Reason*, Kant seems to change tactics. Instead of arguing from our rational agency to our freedom and then to the moral law, Kant begins with the fact of reason, which is understood as our consciousness of the moral law. This of course needs much explanation, and in my third chapter I will consider three influential accounts of this argument, which will help us understand Kant's strategy. But still I argue that these interpretations of Kant's arguments fail to offer accounts that justify morality.

And in my fourth and final chapter, I offer my own account of the fact of reason, which I believe is free of some of the problems I find in the more popular accounts. I do not pretend to offer a definitive account of the fact of reason argument. But I do take myself to be offering a novel one, one that has its own distinct advantages and one that is worth pursuing.

Chapter 2

Morality and Freedom in Groundwork III

2.1 Introduction

In the *Critique of Practical Reason*, Kant tries to articulate a conception of morality that meets the main criteria for an adequate moral theory, which I discussed in the previous chapter. We might also think that he is also trying to articulate a conception free of the problems found in his own *Groundwork*. It is generally accepted that Kant, in his later works, continues to endorse the findings he's made in chapters one and two of the *Groundwork*. The first two chapters of the *Groundwork* are intended to be analytic. That is, Kant wanted to uncover, from our pre-theoretic moral thought, the basic principle of morality that underlies it. He did not intend to revise our ordinary moral thinking, but rather to make it more perspicuous.

One may have doubts about such a methodology. We may grant that common sense opinions count as evidence for any given theory, but they do not, at least for most philosophers, count as decisive evidence. Common sense opinions often turn out to be wrong. The planet, it turns out, is not flat, and motion is relative. And sometimes even common moral opinions turn out to be wrong as well. Even if some of us suspected that slavery was

wrong a long time ago, it was quite common, even among the slaves, to believe that it was acceptable. I here want to defend Kant's methodology, and I offer two responses. The first is that Kant is dealing with fairly abstract features of the nature of morality; in particular, he is concerned with the nature of moral obligation. One notable feature of the ordinary conception of morality, which Kant spends a lot of time developing, is that it seems that we often feel compelled to do what is right even if doing the right action is against our inclinations. That common sense opinions include wrong moral judgments does not imply that they exhibit the wrong idea of morality.

Kant's approach then is to begin by determining the basic abstract features of morality, and then examine the logical consequences of those features. He was doing conceptual analysis on the ordinary conception of morality. If morality is real, it must be close to what ordinary people think it is. But two questions immediately raise themselves: (1) Is he right to think that the basic features he finds in our ordinary conception are really there to be found? He could be wrong about what he finds there. And (2) Why think that an analysis of the concept of morality will reveal anything at all about what is right or wrong? (After all, why think that an analysis of the concept of knowledge will help us understand what is true or false? Can we not come to learn things and acquire knowledge without understanding what it means to know? Do we not in fact so proceed?)

I do not answer the first question, but for the purposes of this project and because I find Kant's account prima facie plausible, I will assume that he is right about many of the basic features of morality. My project after all is to explore how Kant justifies his ethics as being part of practical reason, and it is not to defend Kantian ethics with all its particularities. As for the second question, I believe it is best answered by trying to see how far we get with analysis alone.

2.2 On the Groundwork III Argument

The first two chapters of Kant's *Groundwork* are often treated as the main core of his ethical theory. We are, generally speaking, less concerned with *Groundwork III*. But what is unfortunate about the common treatment of Kant's ethical theory is that it ignores the fact that the first two chapters of the *Groundwork* can only show what morality must look like if it were real. They cannot show that morality is real. For example, if you wanted to disprove the reality of morality (or any other concept), you'd start in roughly the same way. You'd first determine the necessary features of morality. And then secondly, you would show that nothing satisfies or could satisfy all those necessary features. The first two chapters of the *Groundwork* are incomplete on their own; they only "search for" the supreme principle of morality (G 392). Only in chapter 3 does Kant take himself to actually "establish the supreme principle of morality" (G 392). And as a result of his findings in the first two chapters, that task becomes more difficult.

In the first two chapters, Kant begins by showing a crucial feature of the ordinary conception of morality: we are morally obligated to perform certain actions or act in certain ways regardless of whether we want to perform those actions or not. For better or worse, Kant takes this fact at face value. In fact, he tries to show that it is part of the ordinary conception of morality that we take morality to be independent of desire. This is the main point of his discussion regarding the nature of a good will and the concept of duty. He argues that we only actually ascribe moral worth to an action if it is done from the motive of duty. Only if one would perform the morally right action in the absence of any coinciding desire or inclination does the action have moral worth.[1]

[1]Notice that the consequent in this conditional is itself a conditional, but it is a counterfactual one: If a person is acting from a moral motive, then if she were not to have any coinciding desires or inclinations, then she would perform the same action. And so, contrary to common intepretation, Kant does not believe that one must rid oneself of coinciding desires in order to be moral, nor does he believe that we must find out if we are acting from a moral motive.

Kant has now raised the stakes for himself. In order to defend morality, he has to show that a purely moral interest is possible for us. And this is his task in *Groundwork III*. This involves two main things: one, showing that we are free; and two, that we human beings take moral imperatives as indefeasibly normative for us. Unfortunately, those arguments are confusing. Not only do various commentators disagree about how the argument goes, but they disagree about what is being established in *Groundwork III*. Some argue, as Thomas Hill does, that Kant establishes that we are free from a practical point of view, and that is enough. Christine Korsgaard offers a reconstruction of Kant's argument which shows that we must presuppose the Categorical Imperative in order to be free at all. I will consider these arguments in turn. But before I do that, I'd like to mention that the second *Critique* has an argument for the justification of morality which is often thought to be an improvement over the *Groundwork* argument.[2]

As I see it, Kant, in the *Groundwork*, tries to establish the following claims: (1) that our theoretical rationality entails our practical rationality or our freedom, and (2) that our practical rationality or our freedom entails an obligation to the moral law. I will be concerned with trying to understand Kant's arguments for (2) as presented in the *Groundwork*. But in the second *Critique*, Kant says that it is our awareness of morality that reveals our freedom to us, and not the reverse.

I will not offer a full interpretation of the argument in the *Groundwork III*, but consider various problems with it, and why it has difficulties that various commentators have been unable to resolve. I will give a rough sketch of how I think we should try to read *Groundwork III*, but I still argue that the problems persist. And all this will lead into the next chapter, which considers various interpretations of the *Critique* argument which aims

[2]It is not a settled debate whether Kant himself saw the new argument as a replacement for the old argument, he saw the new argument simply as an additional argument, or what he was actually arguing in the *Groundwork*. See "Morality as Freedom" in Korsgaard (1996) for an example of someone who thinks the second *Critique* argument is an improvement over the *Groundwork* argument. See "Kant's Argument for the Rationality of Moral Conduct" in Hill (1992) for someone who believes otherwise.

to avoid the problems found in the *Groundwork.*

2.3 The Setup

I will focus my attention on three arguments in *Groundwork III*. The first is what I call the "Analytic Argument" found in the first two paragraphs of that chapter. Here Kant says that he establishes that "if, therefore, freedom of the will is presupposed, morality together with its principle follows from it by mere analysis of its concept" (G 447). Second, there is the "Under the Idea of Freedom Argument" where Kant asserts that "every being that cannot act otherwise than *under the idea of freedom* is just because of that really free in a practical respect. . . " (G 448). Third, there is the "Intellectual World Argument", where Kant argues that we must take ourselves as "belonging to the intellectual world" in order to explain our interest in the moral law (G 451).

After I give a brief exposition of the Analytic Argument, which consequently sets the stage for the other two arguments, I will consider two influential views on Kant's arguments in the *Groundwork III*. First, I will consider Thomas Hill's account, which exhibits his characteristic virtue of being clear. And second, I will consider an account offered by Christine Korsgaard, whose work has been influential in the revival of Kantian ethics.

2.3.1 The Analytic Argument

The Analytic Argument is contained in two compact paragraphs at the beginning of *Groundwork III*. Unlike the other two arguments, its main claims and its basic order are generally agreed upon. But it remains dense, so I will first try to disentangle its main claims and its basic structure. I will quote at length. Kant writes,

> *Will* is a kind of causality of living beings insofar as they are rational, and

> *freedom* would be that property of such causality that it can be efficient independently of alien causes *determining* it. . .
>
> The preceding definition of freedom is *negative* and therefore unfruitful for its insight into its essence; but there flows from it a *positive* concept of freedom, which is so much the richer and more fruitful. Since the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited, so freedom, although it is not a property of the will in accordance with natural laws, is not for that reason lawless but must instead be a causality in accordance with immutable laws but of a special kind; for otherwise a free will would be an absurdity. Natural necessity was a heteronomy of efficient causes, since every effect was possible only in accordance with the law that something else determines the efficient cause to causality; what, then, can freedom of the will be other than autonomy, that is, the will's property of being a law to itself? But the proposition, the will is in all its actions a law to itself, indicates only the principle, to act on no other maxim than that which can also have as object itself as a universal law. This, however, is precisely the formula of the Categorical Imperative and is the principle of morality; hence a free will and a will under moral laws are one and the same (emphasis all his, G 446–447).

I take Kant to be trying to establish the following claims, each one leading to the next:

> A1: For any being, if it is living and rational, then it also has a will.
>
> A2: For any being, if it is living, rational, and has a will, then it is also negatively free.
>
> A3: For any being, if it is living, rational, has a will and is negatively free, then it is also autonomous. (I here take "autonomy" to mean the same thing as "positive freedom".[3])
>
> A4: For any being, if it is living, rational, has a will, is negatively free, and autonomous, then it takes the Categorical Imperative as its law.

Each of these claims is complex, because each involves concepts that have particular meanings for Kant. Moreover, he provides rather compact arguments for each of these claims, if he provides them at all. Since they are rather compact, I will briefly mention them in this

[3]Many commentators equate the two Kantian concepts, particularly in their understanding of the Analytic Argument. Allison (1990, 95), O'Neill (2000, 42), and Hill (1992, 138–42) all take them to be equivalent, and, I think, with good reason. (As far as I can tell, no one seems to think otherwise.) Kant suggests it himself in the passage quoted above. At its beginning, he writes that positive freedom "flows" from the concept of negative freedom, but his own explication of this refers to "autonomy" rather than positive freedom.

section and expand on them in the subsequent sections when considering the interpretations offered by Thomas Hill and Christine Korsgaard.

Let us first consider A1. It says that for any being, if it is living and rational, then it has a will. Kant does not provide an argument for A1, not in the two paragraphs I quoted above or anywhere else. I do think that an argument for A1 is necessary, and it would have been helpful in understanding exactly what A1 means. We can see why an argument for A1 is in fact needed. The antecedent consists in two parts: being alive and being rational. Being alive does not entail having a will because it is simply not true that all living things have a will. Even if one grants that some non-human animals had wills, it would be difficult to grant them to all animals and absurd to grant them to plants. Also, it is not obvious that being rational necessitates having a will. Kant certainly thought, for instance, that God was rational, but it's less clear whether he thought that God has a will.[4]

Kant gives no direct argument for why the combination of life and rationality in one being would necessitate having a will. Nonetheless, for my purposes here, I will grant that if a being is living and rational, it has a will. After all, my concern here is not with the justification of whether we have wills, but with whether *our* having wills generates an obligation to the moral law. I am primarily concerned with the validity of the Analytic Argument in so far as it pertains to us human beings.

Before discussing A2, I want to say something about what it means for a living, rational being to have a will. Kant says that it is a "kind of causality" and that it can be "efficient", and I think the implication is that a will can act for reasons, which we can think of as a kind of efficient causality, but as something other than natural causality. Because a reason bears a relationship to its intended action very differently from how a cause bears a relationship to its effect, acting for a reason requires conceiving it non-empirically. Consider the following

[4]Kant does have an idea of a holy will, but Nathan Rotenstreich (1985) argues that by Kant's own lights, it is odd to attribute a will to God (38).

thought: I will make dinner tonight, because it is my turn to do so. We cannot say that it is my turn to do so which causes me to make dinner. That would be an affront to the concept of empirical causality; there is no entity or event in space-time corresponding to the terms "my turn" or to the terms "that it is my turn to do so". Only if there were such an entity or event could it be placed in causal relationships with the physical event of my making dinner. But neither "my turn" nor "that it is my turn to make dinner" refer to any particular event.

There may, of course, be a physical correlate to *my belief that it is my turn to make dinner*. But even if it were true that *my belief that it is my turn to do so* causes me to make dinner, that remains irrelevant. Insofar as *my belief that it is my turn to do so* causes me to make dinner, it is not my reason for making dinner. My reason for making dinner is simply that it is my turn to do so; my reason lacks the prefix "my belief that". Insofar as we can regard ourselves as being able to act for reasons, we must also regard ourselves as acting independently of the causal order of the empirical universe. I don't mean to suggest that this implies determinism to be false or that there are events that happen outside of the universe, but rather that we cannot regard acting for reasons in purely empirical terms.[5]

[5]Kant wrote,

> Our blame [of someone who tells a malicious lie] is based on a law of reason whereby we regard reason as a cause that irrespective of all the above-mentioned empirical conditions could have determined, and ought to have determined, the agent to act otherwise. This causality of reason we do not regard as only a co-operating agency, but as complete in itself, even when the sensuous impulses do not favour but are directly opposed to it; the action is ascribed to the agent's intelligible character; in the moment when he utters the lie, the guilt is entirely his. Reason, irrespective of all empirical conditions of the act, is completely free, and the lie is entirely due to its default.
>
> Such imputation clearly shows that we consider reason to be unaffected by these sensible influences, and not liable to alteration. Its appearances—the modes in which it manifests itself in its effects—do alter; but in itself (so we consider) there is no preceding state determining the state that follows. That is to say, it does not belong to the series of sensible conditions which render appearances necessary in accordance with laws of nature. Reason is present in all actions of men at all times and under all circumstances, and is always the same; but it is not itself in time, and does not fall into any new state in which it was not before. (A 555–6/B 583–4)

At this point, it remains open to deny that we actually do act for reasons. We may believe that we act for reasons, and it may even be the case that we must think so, but all these beliefs might turn out to be illusions. Perhaps they are necessary illusions, but they will be illusions nonetheless. Kant admittedly has not yet tried to establish that there are, in fact, rational beings with wills. And I think it is good to keep in mind that Kant tries to establish no such thing within the Analytic Argument: it is merely an analysis of what the implications of there being a rational will are, without thereby supposing that there are rational wills.

Now to look at A2. It says that any living, rational will is negatively free. To say of a will that it is negatively free is to claim, in Kant's words, that it "can be efficient independently of alien causes *determining* it" (G 446, emphasis his). So to say of a rational will that it is negatively free is to say that a rational will can act "independently of alien causes". This has two common interpretations. The first is that Kant means that any rational will that is negatively free can perform actions and that we think of this relationship between the will and its actions in non-empirical terms. This is to say what I have said above about the concept of a rational will—that the relationship between a will and its action cannot be considered as part of the empirical universe.

But I think the second interpretation is better. What Kant likely has in mind by a negatively free will is one that can act independently of the desires and inclinations it happens to have. Here, I take "desires and inclinations" to be the meaning of Kant's phrase "alien causes." Many philosophers, in particular those influenced by Hume, hold that all our actions are functions of our desires and beliefs. This Humean view of human action is sometimes pictured as a sort of quasi-hydraulic system—where desires form the various inputs and beliefs constitute the structure, and together they churn out actions. But for Kant, humans act for reasons; that is, reasons are the basic stuff from which our actions issue as rational agents. Our actions, for Kant, are not the causal result of a set of beliefs enmeshed

with a set of original desires. Rather, for Kant, our actions are 'caused' by our wills, and we perform them for reasons. Desires may sometimes figure into our actions, but they do not ground them. For instance, I desire chocolate, and I choose to satisfy this desire by obtaining some—thereby making my desire for chocolate to constitute part of my reason for so doing. So when Kant writes that we are negatively free, what he means to suggest is that we can act for reasons that do not depend upon the desires and inclinations we happen to have. Given the explanation of the terms "rational will" and "negative freedom", we can now better understand A2.

We can now re-write A2 as follows:

> A2: For any being, if it is living, rational, and has a will, then it can act for reasons that are not dependent on the desires and inclinations it happens to have.

There is an interpretive question that I wish to largely set aside, and that is the question about which terms mean which things. I have interpreted "negative freedom" as the capacity of acting for reasons which do not depend upon desires and inclinations. Others may wish to place the capacity for acting on non-instrumental reasons on the term "autonomy". I set the question aside because I am not so much interested in what Kant intended by his terms but the conceptual move from acting for reasons to acting for reasons that are independent of our inclinations and desires. But one interpretation I do want to reject is that negative freedom is the kind of freedom that is incompatible with determinism. On this interpretation, "alien-causes" means anything empirical, and so being negatively free means that one is free from anything empirical. That would imply that a negatively free action is one that could not be understood in empirical terms—the past and the laws of nature would be insufficient in explaining its occurrence. I don't think Kant intends anything so strong at this point.

It's not clear what Kant's argument for A2 is, nor is it even clear where Kant offers an argument to this effect. One place to look for such an argument is in the section immediately following the Analytic Argument, called "Freedom must be presupposed as a property of the will of all rational beings." But it isn't clear that an argument for A2 can be found there. I will examine Thomas Hill's reasons for thinking that an argument can be found and his explanation of what that argument is.

A3 says that for any being, if is living, rational, has a will, and is negatively free, then it is also autonomous. It is not clear exactly what Kant means by autonomy here. Kant's own terse definition of autonomy is "the property of the will by which it is a law to itself (independently of any property of the objects of volition)" (G 440). I believe that what Kant means is that for a will to be autonomous, two conditions have to be satisfied. First, the will must be able to act in accordance with a normative practical principle whose normative force doesn't depend on the desires and inclinations it happens to have. And second, this normative practical principle must exhibit the agent's independence from the empirical world.

To explain the first, let us recall that what it means for a rational will to be negatively free is that it can act for a reason that does not depend on the desires or inclinations the will happens to have. But to count as acting for a reason, one must appeal, at least implicitly, to a rule of inference which connects the content of one's reason with one's action. We can see how its parallel works in theoretical reason. If I say that the two claims "all dogs go to heaven" and "Fido is a dog" serve as a reason to believe that "Fido will go to heaven", then I appeal to a rule of inference, in this case that of universal instantiation. This is an example of a good reason (provided we have good reason to accept the premises). But we don't always reason well. We often reason badly. For instance, Floyd might argue "All human beings are mortal. Spot is mortal. Therefore, Spot is human." This reason is bad because its general argumentative form is invalid; in other words, Floyd appeals to

a fallacious logical principle. But Floyd is still *reasoning*. What differentiates reasoning from non-reasoning is not whether one is appealing to *valid* logical principles, but whether one is trying to appeal to logical principles at all. Any act of reasoning always implicates the reasoner as accepting its general argumentative form. If the general form is good, then we have a candidate for a good reason. If it is not, then the argument is bad. The point is that reasoning always involves more than just the content of a reason and its conclusion. It also involves a claim about the connection between the two.

As the Kantian thought goes, this is true within practical reason as well. A rational will then requires a normative principle of action. An obvious candidate for a normative principle of practical reason is an instrumental one, which might say that we have a reason to pursue an end, if it helps us satisfy a desire we happen to have. Suppose I desire to open the wine and I know that the corkscrew is a means to open it. I now have a reason (a practical reason) to get the corkscrew. In taking this as a reason, I appeal to a connection between the 'premises' (my desire for wine and my awareness of the corkscrew as a means to open it) and the 'conclusion' (that I shall get the corkscrew)—a connection which might be called the instrumental principle of practical reason. But a negatively free will is one that can act for reasons that do not depend on one's desires or inclinations, and so it will require a normative principle of action that can guide it to action independently of whatever desires it may have. And since an instrumental principle of practical reason requires the presence of antecedently existing desires, a negatively free will would require a non-instrumental principle of practical reason.

The second crucial aspect of autonomy is that this extra normative principle (supposing there is one and only one) must, in some sense, exhibit the agent's autonomy. That is, this non-instrumental principle must somehow come from the rational will itself and not from outside. In sum:

> An agent is autonomous $=_{df}$ (1) an agent can act in accordance with a principle that does not depend on any of her desires, and (2) the principle in question comes from the agent herself.

And so A3 states that any living, negatively free being with a rational will is autonomous. And this leads us into A4, which says that

> A4: For any being, if it is living, rational, has a will, is negatively free, and autonomous, then it takes the Categorical Imperative as its law.

A4 is ambiguous; it is not obvious which of the following Kant intends.

> A4(a): For any being, if it is living, rational, has a will, is negatively free, and autonomous, then it always acts according to the Categorical Imperative.
>
> A4(b) For any being, if it is living, rational, has a will, is negatively free, and autonomous, then it can act according to the Categorical Imperative.
>
> A4(c) For any being, if it is living, rational, has a will, is negatively free, and autonomous, then it recognizes the Categorical Imperative as the law it ought to follow.

I think we can rule out the A4(a) interpretation because Kant believes that human beings satisfy all the conditions, and he knows that human beings do not always act morally. A4(c) might seem problematic, because not all human beings do, in fact, recognize the Categorical Imperative as the law they ought to follow. For an example, you don't have to look further than utilitarians. But I believe that the kind of recognition that Kant would have in mind in A4(c) is a more tacit kind of recognition—much like our recognition of logical laws. For instance, not everyone explicitly acknowledges the principle of non-contradiction, but we might argue that it remains implicit in all our thinking. But following it is arguably a pre-condition for thinking correctly in the first place. On the A4(c) of A4, the Categorical Imperative, as the proposal goes, would have a similar status: it would figure tacitly into all of our practical reasoning, even in the thinking of utilitarians despite their denials. I believe that Kant intends to mean A4(c), for A4(b) seems weak, and it could hardly be the last step

of Kant's Analytic Argument. If A4(b) was the final conclusion, a person might simply regard the Categorical Imperative as optional, and not as something he is compelled to act in accordance with. The demands of morality will not appear to such an agent as demands at all, but merely other courses of possible action. I think Kant intends something stronger than A4(b). But at this point I leave it open whether A4 means A4(b) or A4(c), depending on what the arguments I'll examine can actually establish.

Kant's Analytic Argument tries to show that morality is possible, by deriving it from the more modest claim that we are rational beings with wills. If Kant can show that morality is constitutive of our capacity for practical reason, then there is a lot to give up in denying the possibility of morality.

The other two arguments, the Intellectual World Argument and the Under the Idea of Freedom argument, are more contentious. Various people disagree on how we should even understand whether they present arguments at all. I will examine Hill's and Korsgaard's different but similar reconstructions of the Under the Idea of Freedom argument, and Korsgaard's interpretation of the Intellectual World argument.

2.4 Hill's version

In "Kant's Argument for the Rationality of Moral Conduct," Thomas Hill offers an analysis of the way the arguments in G3 are supposed to run. There is much to learn from Hill's reconstruction, for it makes clear the different interpretations of Kant's claims, and the logical connections between them. I do not think Hill's analysis is fully correct, but we will have a much better understanding of what Kant actually shows and what he can show by following Hill.

As I understand it, Hill sees the Analytic Argument as establishing only A3 and A4(c). According to Hill, a separate argument, what I have called the "Under the Idea of Freedom"

argument, is what establishes A2, which says that any living rational being with a will is negatively free. So if we can establish A2 and if the Analytic Argument can be shown to work, then all Kant has left to do is to show that we human beings can consider ourselves as having rational wills. We are then under the authority of the moral law.

Hill focuses his discussion on Kant's argument from A2 to A3 and the argument for A2, believing that both arguments succeed. Hill is however suspicious of Kant's argument from A3 to A4, and I share his suspicion. But I also believe Kant's argument from A2 to A3 fails, and that the argument for A2, at least as it stands, gets us nowhere. Hill does not, at least in that article, make much of the Intellectual Argument or the Possibility Argument, which appear after the Analytic Argument and the Under the Idea of Freedom Argument. Hill seems to think that that the Analytic Argument and the Under the Idea of Freedom Argument is all that Kant uses to establish A4.

2.4.1 From negative freedom to the moral law

I want to first look at Hill's understanding of the argument from A2 to A3. Let us suppose that A2 is true:

> For any being, if it is living, rational, and has a will, then it is also negatively free.

Above, I argued that what Kant probably means by negative freedom is that a being can act independently of her desires and inclinations. Hill reads it the same way. A3 says the following:

> For any being, if it is living, rational, has a will and is negatively free, then it is also autonomous.

Hill and I have the same interpretation of autonomy: being autonomous means at least two things: (1) that a being can act according to a principle which is not itself based on any of

the agent's particular desires or inclinations, and (2) the principle by which the agent acts comes from the agent herself, and not from some source external to her.[6]

So how does negative freedom imply autonomy? Let us begin with a being who is negatively free. Now, to say of a being with a will that it is negatively free means that the being can act independently of the desires and inclinations she happens to have. But the very idea of a will, according to Kant, is a kind of causality. Here Kant introduces a kind of analogy, and asserts that the causality associated with the will is like natural causality:

> Since the concept of causality brings with it that of laws in accordance with which, by something that we call a cause, something else, namely an effect, must be posited, so freedom, although it is not a property of the will in accordance with natural laws, is not for that reason lawless but must instead be a causality in accordance with immutable laws but of a special kind; for otherwise a free will would be an absurdity. (G 446)

The idea, I take it, is that since free will is a kind of causality, it must behave in accordance with a certain kind of law. For observable events, suppose we say that event A caused event B. On Kant's view, this means that B followed A according to some rule—and this rule will be a general empirical law. Likewise, if we say that a free will causes an event, then there must be some rule by which the free will brings the event about.

I'm not sure I understand this argument. After all, for things that are not empirical events, it isn't clear why we ought to think there is any law of causality which governs their behaviour. I do agree with Kant that the idea of a will producing actions, not according to

[6] Actually, Hill posits that autonomy has four separate requirements rather than two. He writes,

> a will with autonomy is not only negatively free but is committed to at least one principle acknowledged as rational to follow but such that (a) one is not causally determined to accept it or follow it, (b) it does not merely prescribe taking the necessary (or best) means to one's desired ends, (c) the rationality of accepting it does not depend upon contingent facts about what means will serve one's ends or about what ends one happens to desire, and (d) the principle is "one's own" or "given to oneself by oneself," i.e., it expresses a deep commitment from one's "true" nature as a rational and (negatively) free agent. . . (111)

I believe that the two I mention are the most crucial and also the easiest ones to understand. At any rate, adding a third and fourth element would make the argument to A3 even harder to prove.

any law or principle whatsoever, would hardly seem to count as a will, properly speaking. For instance, if my body behaves randomly and gets up to walk across the room against my wishes, it does not seem to be the product of a will (or at least not of my will). Acts of will require some intentions. And if we agree with Kant that intentional action is action for a reason, we can see why an intentional action presupposes an abstract principle—and reasons, as I argued earlier, always presuppose abstract principles.[7]

Suppose that we are negatively free and that we satisfy the first requirement of autonomy—that we can act in accordance with a principle that does not depend on any of our desires. Kant now asks "What, then, can freedom of the will be other than autonomy, that is, the will's property of being a law to itself?" (G 446–7) Kant's conclusion is that the first requirement of autonomy, that the principle not be dependent on one's desires and inclinations, requires the second requirement of autonomy, that the principle comes from within.

Hill thinks that Kant's point here works: that the first requirement of autonomy entails the second. Hill considers two objections, both of which merit discussion. I will consider these objections and then raise a third objection, which shows that Kant's argument cannot succeed.

2.4.1.1 Negative freedom and the meaning of "desire"

The first objection Hill considers is the suggestion that Kant's conclusion contravenes basic views about moral psychology, and in particular it contravenes the Humean view of motivation. Kant's thesis of negative freedom states that a rational agent with a will can act independently of her desires, whereas the Humean view states that any action requires both a belief and a desire. A popular formulation of the Humean theory of motivation is Michael Smith's.

[7]Not all philosophers believe that intentional actions always require a reason. For instance Rosalind Hursthouse (1994) argues that a reason isn't always required.

> A reason R at time t constitutes a motivating reason of an agent A to F iff there is a G such that R at t consists of a desire of A to G and a belief that were he to F he would G, where beliefs and desires are distinct existences. (Smith 1987, 36)

This is open to several interpretations, depending mostly on what is meant by a desire. Given what he says, Hill might accept a trivial reading of it given by a rather wide and trivial understanding of "desire" such that the Humean thesis does not contradict Kant's thesis of negative freedom.[8]

Hill believes that Kant's use of the words "desires and inclinations" in his formulation of negative freedom is one of three possible uses. (Since the phrase "desires and inclinations" is cumbersome, I shall simply use the word "desires" instead.) Let us distinguish between the wider, the narrower, and the empirically discernible senses of the word "desire", the last of which Hill thinks Kant intends.

> Wider: An agent A has a desire to F if and only if A is motivated to F. (Hill 1985, 113)

According to Hill, on this reading, our only way of knowing whether an agent desired to do something is whether he in fact did it, provided that he had other choices. On this reading of "desire", Kant's thesis of negative freedom will turn out to be trivially false. To say that we can act independently of our desires, on the wider reading of "desire", will mean that we can F independently of being motivated to F. But since Kant's eventual point is not that we can act morally independently of being motivated to act morally, we must reject this wider sense of desire.

There is also another common reading of desire, which Hill calls the narrower conception.

[8]Hill never actually considers the Humean theory as I've stated it or Michael Smith's view. Rather, he wants to grant that there is a sense in which it is always true that one cannot act independently of one's desires. That is, motivations to act always require desires.

Narrower: An agent A has a desire to F if and only if A feels a desire to F.

On this narrower reading, Kant's thesis that we can act independently of desires turns out to be quite plausible. We can certainly and often do act in ways that do not depend on any felt desire or inclination. For example, I choose to continue running instead of quitting though the muscles in my legs tell me to stop. Common sense tells us that we have desires that we do not currently feel. I have a desire that my friends do well, and I may often feel it, but it would be wrong to say that I lose that desire in those moments I do not feel it. And we most certainly act on these unfelt desires, despite what desires we may be feeling. I may choose to finish the marathon because I have that desire even if I can't feel that desire. Hume, for instance, acknowledged the existence of calm passions, which move us to act but may not be felt, and he believed that we often confused our calm passions with principles (T 417). According to Hill, Kant must mean something by desire such that his thesis of negative freedom is neither trivially false (as on the wider reading) or obviously true (as on the narrow reading). Hill rightly rejects this narrower understanding of the word "desire" as what Kant means.

Hill offers his own interpretation of what Kant must mean by "desire".

> An agent A has a desire to F if and only if A has an empirically discernible motivation to F, and it is not something A has in virtue of being a rational agent as such. (Hill 1987, 114)

He says that the term is meant to "include desires in the narrow sense but also Hume's 'calm passions' and any other preference, liking, aversion, love, hate, etc. which any agents might lack and which is not attributed solely because they acted voluntarily" (114). Hill's point seems to be that we ought to think of our desires as contingent—that we may or may not be so motivated to satisfy them—because Kant's point seems to be with whether we have any *necessary* motivations, ones that arise in virtue of our being rational agents.

But I have two issues with Hill's account. First, this account of desire smuggles in an undefended theory about human motivation. In particular, it offers two ways of dividing all possible motivating reasons, and it assumes that the two ways are equivalent. He first divides those motivations that arise in virtue of our being rational agents from those that do not. And then he divides those we have necessarily from those we might lack. It is however unclear to me that those that *can* arise in virtue of our being rational agents *will* necessarily arise in virtue of our being rational agents. Hill's definition precludes the possibility that some rational agents may simply never come to be motivated by certain rational constraints, and so it is hardly theory neutral. This is easy to see in theoretical reason. In virtue of our being rational, we can all come to learn that there are an infinite number of prime numbers, but we are not all necessarily compelled to believe it. In practical reason, we can imagine that in virtue of our rational agency, we are committed to the Formula of the Kingdom of Ends, for instance, but at the same time, we can imagine many of our rational agents who are unmotivated by it. And on the other side, Kant believes that as human beings we all necessarily desire happiness, but he nonetheless maintains that such a desire does not arise from our being rational agents.

The second issue is this. Suppose we have a motivation to do F, and we are wondering whether it is a desire in Hill's sense of the word. All we can do is ask whether we could have it in virtue of our being rational agents or whether we have that motivation necessarily. If we were motivated to act morally, for instance, and we wanted to know whether it was the result of a desire, we would never know for sure. All we could do is ask if we *could* have it in virtue of our being rational agents alone, and if we discover that we could, then it would not be a desire in Hill's sense of the word. (How would we know if we *could* have a motivation in virtue of being rational agents? One way might be to see if it follows from the concept of rationality alone.) But all this leads to the troubling result: we could not, in principle, have a desire to do anything that is required of us in virtue of our rational agency.

A definition of desire shouldn't be able to rule this out from the start.

But I think Hill is driving at something right. The point about desires and rationality is meant to be about the possible *sources* of motivation—about whether they come from our beliefs alone or from some other source. So I will offer a partial account of what I think Kant intends by "desire". Let us go back to Humean theory of motivation.

> Humean theory: A reason R at time t constitutes a motivating reason of an agent A to F iff there is a G such that R at t consists of a desire of A to G and a belief that were he to F he would G, where beliefs and desires are distinct existences.

The Humean theory requires that beliefs and desires be distinct existences, and I think this requirement is central to how we understand what is meant by desires. So let me offer a criterion of desire.

> An agent A has a desire to F only if there is something distinct from any of A's beliefs, which motivates A to F, and no subset of A's beliefs necessitates A being motivated to F.

Understood this way, there can be no contradiction between any set of beliefs and any set of desires, which is how I understand the claim that beliefs and desires are distinct existences.[9] My partial definition doesn't give much insight as to what a desire is, but that is not really needed for my purposes.

My intention is to offer some understanding of Kant's thesis of negative freedom such that is neither trivially false or obviously true, and is in line with much of what Kant says. Now under my interpretation, to say that a rational agent can act independently of her desires is to say that an agent can act on her beliefs without the aid of extra motivation. That is, Kant's thesis of negative freedom amounts to a version of internalism:

[9] According to Michael Smith (1995), to say that beliefs and desires are distinct existences is to say that "it is always at least possible for agents who are in some particular belief-like state not to be in some particular desire-like state; that the two can always be pulled apart, at least modally" (119).

> *Internalism*: Having certain beliefs is deeply connected with being motivated to act in accordance with those beliefs.[10]

And eventually, Kant will argue for a version of *moral internalism*.

> *Moral internalism*: Having moral beliefs is deeply connected with being motivated.[11]

This section started off by examining an objection to the thesis of negative freedom, which stated that it contravenes basic principles of psychology, and in particular the Humean view. I concede that the thesis of negative freedom contradicts the Humean theory of motivation, but this isn't necessarily a problem. The Humean theory is more than just the view that desires are necessary for motivation, but that desires are independent of beliefs. This theory is not trivially true, and so it's not a great loss that Kant's thesis of negative of freedom contradicts it.

2.4.1.2 The Hypothetical Imperative

So we have been looking at the move in Kant's argument that says satisfying the first requirement of autonomy entails satisfying the second—that being able to act in accordance with a principle that doesn't depend on one's desires implies that this principle must come from the agent herself. Hill considers a second objection to this move (Hill, 114). I think it is an important objection, and my answer to the objection differs radically from Hill's. Recall that Kant's Analytic Argument begins with an assumption of negative freedom, from which he derives first aspect autonomy and then the moral law. Autonomy, as I said, has at least two parts.

[10] Internalism is usually understood to state something less strong, something more akin to "having certain *judgments* are necessarily connected with being motivated. But since Kant is a cognitivist, (that is, he believes that moral *judgments* express *beliefs* capable of being true or false) I do not think that there is much harm in my reformulation.

[11] In the preceding chapter, I defined moral internalism differently. I do not take these differences trivially. But for my purposes in this context, my definition here will suffice.

> An agent is autonomous $=_{df}$ (1) an agent can act in accordance with a principle that does not depend on any of the agent's desires, and (2) the principle in question comes from the agent herself.

In order for Kant to show that a negatively free being is also autonomous, he has to show that both aspects of autonomy are implied. I have argued, along with Kant and Hill, that the first aspect of autonomy follows from the thesis of negative freedom. But unlike them, I am unconvinced that the second aspect comes on the heels of the first. They also say that the only possible candidate for such a principle is the first formulation of the Categorical Imperative. If all this is right, Kant has arrived at his desired conclusion, A4(c), that all rational beings are obliged to follow the first formulation of the Categorical Imperative (the formula of universal law or "FUL").

The objection is this: aren't there principles of practical reason other than the Categorical Imperative? And if there are, then something must have gone wrong in Kant's argument. Either the first requirement of autonomy fails to entail the second, or autonomy doesn't lead to the Categorical Imperative.

This objection becomes weightier once we can articulate a principle, other than the Categorical Imperative, which satisfies the first requirement of autonomy. And there is an obvious candidate: the general principle of instrumental reason.

> An agent ought to either take the necessary means to a desire (or chosen end) or give up the desire (or end).

Many take Kant to endorse a version of this principle called the Hypothetical Imperative (capital "H", capital "I").[12] It says the following:

> An agent ought to either will the necessary means to a willed end or give up willing that end.

[12] See Hill's "The Hypothetical Imperative" (reprinted in 1992), Korsgaard (1997), Herman (1993: 144, 214-15).

For Kant, willing something is more than just desiring it; it involves focusing one's efforts on pursuing it. The Hypothetical Imperative, in combination with what one wills, issues particular hypothetical imperatives. But we don't actually need to will anything at all for the Hypothetical Imperative to be valid for us. It mentions willed ends, but it doesn't mention any particular end that any particular agent must will. We are required to follow it regardless of whether we will anything or not. In other words, the Hypothetical Imperative turns out to be a kind of categorical imperative.

The Hypothetical Imperative may be vacuous for a man who wills nothing, but that doesn't mean that he isn't obliged to it. Many of our moral obligations are also vacuous. For instance, "If you have prisoners of war, you must feed them" is valid for me, even though I don't have any prisoners of war. And so in addition to all the *moral* obligations I have, it turns out that I also have this obligation:

> Will the necessary means to a willed end or give up willing that end.

This principle is valid for me regardless of whatever I will, and even of whether I will anything at all. So the first requirement of autonomy is satisfied, but that doesn't lead us automatically to the Categorical Imperative.

But does this show that the first requirement of autonomy doesn't lead to the second? Or does it show that autonomy (with both aspects) doesn't lead to the Categorical Imperative? It depends on whether you think the Hypothetical Imperative satisfies the second requirement of autonomy. That is, if you believe that the Hypothetical Imperative comes from the agent herself, then this objection shows that autonomy doesn't lead to the Categorical Imperative. And if you don't, then this objection shows that the first requirement of autonomy doesn't lead to the second. In other words, it doesn't really matter. The objection works regardless. In short, there is a principle, the Hypothetical Imperative, other than the Categorical Imperative that seems to satisfy the first requirement of autonomy. How is it

ruled out?

Hill wants to rule out the Hypothetical Imperative on the grounds that it is a weak normative principle. It may never tell us that we ought to do anything in particular. This is true. Even if we willed ends, which issue particular hypothetical imperatives, we can always just give up on willing them. Hill says that it would only ever give us "an option rather than a course of action" (115). Applying the Hypothetical Imperative to a particular willed end results in "I ought to will the necessary means to my willed end or give up willing that end." And if the Hypothetical Imperative was the only principle of practical rationality, there would never be anything in particular we would have to do. We would only ever get disjunctive requests. Either do X or give up Y. So, according to Hill, in order for practical rationality to issue distinct non-disjunctive demands, we therefore need more than just the Hypothetical Imperative.

There are two obvious difficulties with this reply. First, this absurdum, a world where the Hypothetical Imperative is the only practical principle, may not be as disastrous as Hill fears. Maybe we aren't required to do anything in virtue of our rationality at all. Pure practical reason might not be substantive at all. Moral sceptics, Humeans, and a host of many other philosophers alike are perfectly willing to accept such a result.

Hill's reply also faces a second problem; this time it is a textual problem. Kant says nothing like what Hill says in the Analytic Argument or anywhere else. He never says that a conception of rational will requires more than just a Hypothetical Imperative because it would otherwise be ineffectual, and so we thus require a further principle: the Categorical Imperative. Rather, the principle of morality is supposed to follow from freedom of the will "by mere analysis of its concept" (G 447).

And unless we can find a further reason which shows that the rational will is more substantive than what has so far been presumed—that it is negatively free and can follow a principle that does not depend on any particular desire or inclination—Hill's interpretation

will depend upon importing further assumptions. Hill has to show that it is part of the very concept of a rational will, and not just because it would otherwise make a rational will rather ineffectual, that a more substantive moral principle can be found. But this is unlikely, because this is what the Analytic Argument was intended to show in the first place.

Fortunately, there is a better and simpler reason why this objection fails, namely, that Kant does not believe in the Hypothetical Imperative. I argue that Kant does not endorse any normative principle of the following form.

> An agent ought either to will the necessary means to a willed end or give up willing that end.

I grant that Kant held that we have particular hypothetical imperatives that enjoin us to pursue particular means given particular willed ends. But we have no obligation to any overarching principle called the Hypothetical Imperative.

There is of course a general principle that underlies the particular hypothetical imperatives, but it is not a principle of conduct or an imperative that we must accept, as Hill believes.[13] Rather it is more akin to a rule of inference, except it tells us what we ought to do, rather than what we ought to believe, provided other things we will.[14] First, Hill provides little textual evidence for thinking that Kant believed in the Hypothetical imperative. Kant consistently refers to "hypothetical imperatives" rather than to "the Hypothetical Imperative". The one passage Hill does cite which might indicate that Kant believes in a singular Hypothetical Imperative is this: "...[T]he imperative that commands volition of

[13]Mark Schroeder (2005) also makes the case that there is no Hypothetical Imperative. Schroeder argues that we ought to reject a wide scope interpretation of the hypothetical imperative in favour of a consequent scope interpretation. That is, we ought to reject "$\forall x \forall e$ (x ought (if x wills end e, then x takes the means to e)" in favour of "$\forall x \forall e$ (if x wills end e, then x ought (x takes the means to e)) as an interpretation of the general form of hypothetical imperatives. I'm not entirely convinced by his consequent scope interpretation, but I am convinced by his reasons for rejecting the wide scope interpretation. My arguments against the wide scope reading are distinct from his, and my own positive account is decidedly less clear.

[14]The distinction between a rule of inference and a premise is crucial, but it is not always easy to disentangle them. Lewis Carroll (1895) famously emphasized the importance of the distinction.

the means for him who wills the end…" (G 419). But Hill takes this out of context. The full passage is this:

> This imperative of prudence would, nevertheless, be an analytic practical proposition if it is supposed that the means to happiness can be assigned with certainty; for it is distinguished from the imperative of skill only in this: that in the case of the latter end is merely possible, whereas in the former it is given; but since both merely command the means to which it is presupposed one wills as an end, *the imperative that commands volition of the means for him who wills the end* is in both cases analytic. Hence there is no difficulty with respect to the possibility of such an imperative. (G 419, emphasis mine.)

Read properly, this passage does not refer to any singular Hypothetical Imperative. Rather Kant is talking about the particular hypothetical imperative, "in both cases" of prudence and skill. It does not refer to an imperative that is common to both cases of prudence and skill. Kant means to say that the particular hypothetical imperative, in the case of prudence (if we knew the certain means to happiness) is analytic, and the particular hypothetical imperative, in the case of a skill, is also analytic. There is no implication that there is one Hypothetical Imperative. At most, we can read that there is a similarity to the imperatives in both the prudential and the skill case.

And what I think is common to both the prudential case and the skill case is that they are both hypothetical imperatives. The general form of a hypothetical imperative is not itself an imperative at all. It is not a principle of conduct to acknowledge and accept, but as a principle that tells us what we ought to do given that we will particular ends. In Kant's words:

> This proposition is, as regards the volition, analytic; for in the volition of an object as my effect, my causality as acting cause, that is, the use of means, is already thought, and the imperative extracts the concept of actions necessary to this end merely from the concept of a volition of this end. (G 417)

The principle in question is this:

> Whoever wills the end also wills (insofar as reason has decisive influence on his actions) the indispensably necessary means to it that are within his power. (G 417)

Let us take Kant to mean the following:

> If an agent wills an end, then she wills the necessary means to that end.

We may think that such a claim is too strong, as it would seem to preclude any case of an agent failing to will the means when he wills the ends. In other words, it seems to make cases of weakness of will impossible. But I'm unconvinced. If we keep in mind that "willing" for Kant means focusing and directing one's concerted efforts at achieving a certain end, and that it is more than just desiring an end, choosing it, or even intending it, the claim no longer looks to be too strong. Examples of weak wills do not concern "willing" in Kant's sense of the word.

Kant provides the clause "insofar as reason has decisive influence on his actions", and many philosophers seem to think that Kant means "If an agent wills an end, then *either* she wills the necessary means to that end *or* she is not being fully rational" which is somehow equivalent to this claim "If an agent wills an end, then she *ought* to will the necessary means." I'm uncertain if this is the right way to think about the hypothetical imperatives, but there has to be a way of getting particular hypothetical imperatives to result. Kant does believe that it is constitutive of willing an end that one is obligated to will the necessary means to that end. And the paraphrase "if an agent wills an end, then she ought to will the necessary means" captures that thought.

No agent has any requirement to satisfy or meet this principle, because it isn't a general normative principle of practical rationality. It describes and explicates what it means to will; and we have obligations only if we will ends.

I will provide one further textual reason for denying that Kant posited any Hypothetical Imperative. Kant says that happiness is an indeterminate concept, but that if it were

determinate, the imperative of prudence would be analytic (G 419). I will argue that what Kant says does not make sense if we suppose there to be a Hypothetical Imperative. Depending on whether there is a Hypothetical Imperative, which is itself an imperative and issues particular hypothetical imperatives, or only hypothetical imperatives, the imperative of prudence would be different in each case. Suppose, for a second, that there is something called the Hypothetical Imperative, and it is analytic as Hill and others claim it to be. It says,

> An agent ought either to take the necessary means to what one wills or give up willing that end.

The supposed imperative of prudence would then be:

> An agent ought either to take the necessary means to happiness or give up happiness.

But this imperative would be analytically true *independently* of whether happiness is a determinate concept or not. In fact, given the schematic quality of the Hypothetical Imperative and its own supposed analyticity, any instantiation of it (even if what one considers to will are impossible) will also be analytic.

On my reading, there is no Hypothetical Imperative that we ought to follow. What we have instead is this:

> If an agent wills an end, then she wills the necessary means to that end.

Understood this way, we can see why Kant thinks happiness being an indeterminate concept would preclude the imperative of prudence from being analytic. Kant says that happiness is too indeterminate a concept, which I take to mean that there are no necessary and sufficient means to happiness. Since one cannot automatically take the necessary and

sufficient means to an end when there aren't any, it doesn't follow that whoever wills happiness automatically wills the necessary means to it. On the other hand, if happiness were a determinate concept, and say, the set (c, d, e) are the necessary and sufficient means to happiness and suppose that we know that, then we would focus our efforts in achieving $c, d,$ and e. And this imperative would follow from an analysis of willing happiness. It would, as a result, be analytic for us that we ought to do c, d, and e. But the only "imperative" of prudence, for Kant, is hardly an imperative at all. There are only "counsels of prudence", which are more uncertain and function more as general pieces of advice (G 416).[15]

Kant writes, "If only it were as easy to give a determinate concept of happiness, imperatives of prudence would agree entirely with those of skill and would be just as analytic" (G 417). Given that happiness is not a determinate concept, it makes sense, on my reading, for Kant to deny that prudence can issue commands. It makes little sense on Hill's reading, because on Hill's reading, there is still a principle that one ought to obey: "take the necessary means to what one wills or give up willing that end".

I want to offer another reason to reject the idea that Kant believed in the Hypothetical Imperative: doing so would make more sense of Kant's Analytic Argument. The original objection, if you'll recall, is this: there seems to be a principle, other than the Categorical Imperative, which satisfies the first requirement of autonomy, namely the Hypothetical Imperative. If there is no Hypothetical Imperative, the objection weakens. And if, on the other hand, Kant were to have held that there was a Hypothetical Imperative, he himself would have recognized that there was a huge gap between a negatively free rational will, one that can act independently of any particular desires or inclinations, and acting autonomously according to the Categorical Imperative.

[15] Kant says that the counsels of prudence "involve necessity" (G 416), but at the same time, they cannot "command at all" (G 418). Paton (1967) explains this disparity by saying that the "counsels of prudence are... binding, since it is mere folly to wreck one's happiness, an end which is very far from being arbitrarily chosen" (116). But that they are not commands because it is "impossible to be sure wherein a particular individual will find his happiness" (115).

I now want to raise a third objection against Kant's argument from negative freedom to autonomy. I said above that on Kant's view an agent is autonomous if and only if (1) the agent can act in accordance with a principle that does not depend on any of the agent's desires, and (2) the principle in question comes from the agent herself. Kant seems to think that (1) automatically leads into to (2). He asks rhetorically, "what, then, can freedom of the will be other than autonomy, that is the will's property of being a law to itself?" (G 446) But I don't see why (1) automatically implies (2). The rough idea of why (1) and (2) go together was that any supposedly external authority, such as God or the King or the legislature, only has authority because it depends on desires and inclinations of the subjects, for instance, fear of punishment, desire to be loyal, and so on. All these external authorities, if they have any authority for an agent, must depend on one's desires and inclinations. The falsity of (2) would thus seem to require the falsity of (1); and thus the first aspect of autonomy implies the second.

The difficulty with this idea is that not only God, the King, or the legislature count as an external authority. If there are external facts about the world as at is, there might also be external facts about what the world *ought* to be like. For instance, rational intuitionists such as Moore, Ross, and Prichard believe that there are normative truths, about the way the world ought to be, which are external to us. And moreover, the truthmakers of these normative claims do not depend on the particular desires or inclinations a person happens to have. Moral truths are out there to be found. Additionally, Moore, Ross, and Prichard all believe that we can be motivated to act according to these moral truths, even though we know full well that they may not serve any of our particular desires or inclinations (Prichard 30). Hence, rational intuitionism seems to suggest it is a real possibility that the first aspect of autonomy be satisfied without the second being satisfied. That is, we can act according to principles which are not based on the desires or inclinations we happen to have, and yet these principles remain external to us.

The two requirements of autonomy don't obviously imply one another. According to the first requirement, an agent must be able to act according to a principle that does not depend on the particular inclinations or desires she happens to have; this is a claim about practical reason. Trying to develop a theory that meets this demand is trying to place our ethical requirements within a broader conception of practical reason. The second requirement, that the principle of action come from the agent herself and not from outside, seems to be a metaethical claim about the source of the principle, or about whatever it is that makes it true. But it is plausible, and rather natural, to think that what motivates us to follow a certain principle is different from what makes that principle valid. For instance, my motivation for following the law might be that I fear punishment, but it isn't my fear of punishment, or any other contingent feature of me, which makes the law valid. (For a law to be valid, it may be the case that violations of it must be punishable. But that is a different matter.) Likewise, it may be non-natural features of the external world which make certain actions right, but the motivation for following such actions do not also have to come from the external world. I may always have a reason to do what is moral, yet what it is that makes an act moral may be something external to me. Kant appears to be assuming that if external things make a normative principle valid, then it must also be external things which motivate the person to follow that principle.

Even though I have been focusing almost exclusively on how negative freedom entails autonomy (or the step from A2 to A3), my general concern is with the whole of Kant's argument, in particular the step from rational wills being negatively free to being obligated to act in accordance with the Categorical Imperative. But there is a further problem.

It isn't clear that Kant actually offers any argument for the thesis that rational agents are negatively free nor is it clear where he does. Thomas Hill believes that Kant does offer an argument, which can be found in the section immediately following the passages containing the Analytic Argument. But Onora O'Neill doesn't think that this passage indicates much

of anything, let alone an argument for our negative freedom (1989, 55). Here is the passage:

> . . . since [morality] must be derived solely from the property of freedom, freedom must also be proved as a property of all rational beings; and it is not enough to demonstrate it from certain supposed experiences of human nature (though this is also absolutely impossible and it can be demonstrated only a priori), but it must be proved as belonging to the activity of all beings whatever that are rational and endowed with a will. I say now: every being that cannot act otherwise than under the idea of freedom is just because of that really free in a practical respect, that is, all laws that are inseparably bound up with freedom hold for him just as if his will had been validly pronounced also in itself and in theoretical philosophy. Now I assert that to every rational being with a will we must necessarily lend the idea of freedom also, under which alone he can acts. (G 448)

Here we see that Kant first says that we need to prove that rational wills are free, and that such a proof must be done a priori. And then Kant proceeds to make claims about how we must regard beings who act "under the idea of freedom" as "really free in a practical respect". Kant's claims here are difficult to understand. Suppose when we act, we must regard ourselves as free, in other words, as if we are free to choose one of our options before us. I take it this is the meaning of acting under the idea of freedom. And this means that we are free in a practical respect. But it's hard to see what it means to say that "we are free in a practical respect" apart from thinking that when we make decisions, we make them as if we were free to choose one of our options before us. In other words, it's hard to see how being "free in a practical respect" means something more than acting "under the idea of freedom". We can see why O'Neill was sceptical of whether there was any argument at all (1989, 55). And as an interpretation of A2, the second claim in the Analytic Argument, the textual evidence is slim. The passage above mentions "freedom", but it says nothing about "negative" or "positive" aspects of freedom.

Nonetheless Hill thinks an argument can be found. Its main claims are as follows.

U1: A rational will cannot act except under the Idea of freedom.

U2: Any being that cannot act except under the Idea of freedom is free from a practical point of view.

U3: Therefore, rational wills are free from a practical point of view. (Hill 116)

The basic idea is this: all rational beings see themselves as negatively free when they act or deliberate. That is, they necessarily see themselves as capable of acting for reasons that do not depend upon the particular desires or inclinations they happen to have. And so U1: every rational being with a will cannot act except under the idea of freedom. U2 says that any being that cannot act except under the idea of freedom is really free in a practical respect.

Hill takes the argument to establish the legitimacy of a qualification "from a practical point of view". Hill thinks that its unstated conclusion is this:

> "Free from a practical point of view" is sufficient for purposes of the rest of the arguments for the rationality of moral conduct; and so for the purposes of that argument the qualification can be dropped. (117)

Hill's idea is this qualification—from a practical point of view—is meant to constrain the results of the Analytic Argument, making Kant's conclusion more palatable. All the controversial conclusions, no matter how seemingly metaethical and metaphysical, are qualified by being under the assumption of a practical perspective. The Analytic Argument is now supposedly understood in the following way:

A1: For any being, if it is living and rational, then it also has a will.

A2': From a practical point of view [for any being, if it is living, rational, and has a will, then it is also negatively free.]

A3': From a practical point of view [for any being, if it is living, rational, has a will and is negatively free, then it is also autonomous.]

A4': From a practical point of view [for any being, if it is living, rational, has a will, is negatively free, and autonomous, then it takes the Categorical Imperative as its law.]

I think this line of reasoning is plausible, but it requires textual stretching. Almost nothing Kant says implies that we should regard A3 or A4 as qualified by "from a practical point of view".

Also, the occurrences of "from a practical point of view" above function as a sentential operator, but unfortunately for Hill, not all sentential operators have the necessary closure property, without which the argument suffers from additional difficulties. Here's what I mean. It is not the case that for all sentential operators O, if $O[A]$ and $A \rightarrow B$, then $O[B]$. An obvious example is the sentential operator "not"; $\{\sim A,\ A \rightarrow B\}$ does not entail $\sim B$. Other examples of sentential operators which do not have closure include "S knows that", "S believes that", and other intentional contexts. For example, just because John believes that Muhammad Ali is a boxer, it doesn't follow that John believes that Cassius Clay is a boxer, even though Muhammad Ali is Cassius Clay. Some sentential operators are, of course, closed, but Hill must give a reason to think that "in a practical reason" does obey the relevant closure principle. So just because A2 implies A3, it doesn't follow that "In a practical respect, A2" implies "In a practical respect, A3". We need to know more about the sentential operator "In a practical respect" before we allow ourselves to make such inferences. We must regard Hill's reconstruction of Kant's argument in G3 as remaining invalid. Putting an "in a practical respect" to qualify one's claims can preclude rather than help draw the desired conclusions. Hence, Hill's suggestion doesn't seem to do what he wants.

Though Hill does not directly consider my worry, he makes a remark which might offer a way out. He writes that rational agents "should accept, for all practical purposes, that they are free in this sense. That is, they should accept any implication that 'as a rational agent I am (negatively) free' as a reasonable assumption in all their deliberations about what to do" (119). The idea is that we should understand the Analytic Argument as addressed to the rational agent herself. Let us make an analogy. Consider the sentential operator "I

know that". Suppose that $P \rightarrow Q$ and that I know that P. It doesn't follow that I know that Q, because I may not yet believe that $P \rightarrow Q$. But if I'm shown that $P \rightarrow Q$, then I *should* accept Q (or reject P). So the idea is that the rational agent herself should accept the implications of whatever she herself already believes or must believe. So, if she must believe that she is negatively free, then she should accept the implications of that negative freedom.

I believe that Hill's suggestion is rather attractive, and I agree that from our own perspective, we might have to accept the implications of our thoughts. But I think that this approach raises more problems than it solves. It makes the process too internal. First, an agent should only accept the implications of her beliefs if she *believes* the implications. If she doesn't *believe* that negative freedom implies autonomy, then she doesn't have reason to think that she is autonomous. We, as external observers, perhaps convinced that negative freedom does imply autonomy, may think that she is autonomous, but that would be irrelevant, because we, as external observers haven't been shown that she is, in fact, negatively free. Second, this would make the move to A4 and Kant's overall conclusion quite the leap. It would only be rational agents who actually believed that negative freedom implied autonomy who are under the moral law. But Kant's conclusion, the one he needs, is that the moral law applies to all rational beings with a will, not just to those who *believe* that negative freedom implies autonomy and the moral law. Kant's argument must be stronger than what Hill's interpretation can possibly offer. People can be held responsible for not acting morally. That is a part of Kant's, and ordinary people's, conception of morality.

2.5 Korsgaard's Reconstruction

Korsgaard doesn't provide any direct interpretation of the Analytic Argument, as I have called it. The closest she comes is in a paper called "Morality as Freedom" (Reprinted in

Korsgaard (1996)). In it, she presents an interesting argument for the Categorical Imperative, which I suspect might be the right interpretation of Kant's argument in the *Groundwork*, but I am sceptical of its success as an argument. But if the argument succeeds, then it can show how being negatively free implies being committed to the Categorical Imperative. And this would complete the gap in the Analytic Argument. She also addresses the Intellectual World argument, which I will discuss afterwards.

2.5.1 The Regress Argument

Korsgaard provides an argument which aims to show that the Categorical Imperative must be assumed in order for intentional action to be possible at all. She tries to show that the only principle which could be used to underlie a reason for an action is the Categorical Imperative; all other principles fail either on pain of being inadequate to the task or on pain of regress.

We have to first try to understand what is involved in performing an action for a reason. Let us consider an action A performed by S for a particular reason R. According to Korsgaard, in order for an agent to regard R as a reason for an action, at the very least, he has to believe that R fits or satisfies a principle of reason which allows him to consider R to be a reason for an action in the first place. There is nothing particularly controversial about this. In order to perform an action for a reason, we must believe that there is a connection between one of our beliefs and our proposed action. For instance, we cannot believe that roller coasters are fun and take that as the reason for going on a roller coaster without thinking that a prima facie reason for doing something is its potential for fun. Otherwise, the mere belief that roller coasters are fun would be merely one of many beliefs we had, and it only serves to function as a reason for action because we have the additional belief that going on a roller coaster helps us get what we want, which in this case is fun. In other

words, we need to appeal, at least implicitly, to an abstract principle that makes our beliefs relevant to action in order for them to serve as reasons for action.

Korsgaard expresses the abstract principle which relates our beliefs and actions as follows:

> I will do this action, in order to get what I desire. (164)

But according to Korsgaard, this principle needs additional support. In other words, she thinks that despite having the beliefs that roller coasters are fun and that going on a roller coaster will help us get what we want (fun), she nonetheless holds that we would be unmotivated to go on a roller coaster unless we adopted a further principle, which she expresses as follows:

> I will make it my end to have the things I desire. (164)

In contrast, Humeans about instrumental rationality will think that the first principle suffices, or at least some analog of it does, and that there is no need for further abstract principles of practical reason. They hold that a motivated action by an agent requires both beliefs and desires, and that beliefs and desires are distinct. We'll recall that Michael Smith characterizes the Humean thesis of motivating reasons for action as follows:

> A reason R at time t constitutes a motivating reason of an agent A to F iff there is a G such that R at t consists of a desire of A to G and a belief that were he to F he would G, where beliefs and desires are distinct existences. (Smith 1987, 36)

This implies that, on the Humean view, so long as the agent A desires to G and believes F leads to G, then that suffices to provide a motivating reason for her to F (provided the condition that beliefs and desires are distinct existences is satisfied).

Korsgaard tries to undercut the Humean view of practical reason by suggesting that what it offers in the way of a practical reason remains insufficient. Her argument takes the

form of a regress argument. When it comes to justifications, a regress can loom almost anywhere. For instance, suppose you are at a restaurant and chose to have the salad. You come to this choice because you have these two further beliefs: that the salad is healthier than the twice-fried chicken, which happens to be the only other option; and that you desire to be healthy. And so there are two obvious regress type questions one can ask here: (i) why do you believe that salad is healthier? (ii) why do you desire to be healthy? Being unable to give an adequate justification to either of these might undermine your decision.

But Korsgaard wants to ask a separate regress question, "Why should you try to satisfy your desires?" (164) So you desire to be healthy, you believe that salad is the healthier option, and let us suppose that you also believe that choosing the salad will help satisfy your desire. You think that those three facts justify choosing the salad. And so you thus choose. But according to Korsgaard, this would not be sufficient justification unless you took it as a general principle to try to satisfy your desires. If, on the other hand, you didn't, then those three beliefs of yours would strike you only as facts. If instead you took it as a general principle to try to stifle your desires and leave them unsatisfied, then the sum of your belief that salad is the healthier choice, your desire to be healthy, and the application of the instrumental principle (which would mean that choosing the salad satisfies your desire) would altogether remain unmotivating. In short, Korsgaard believes that you need to be able explain why we should try to satisfy our desires in order for those beliefs to be motivating.

So what justification could we give to thinking that we should try to satisfy our desires? In other words, why should we each adopt the principle stated above "I will make it my end to have the things I desire"? Korsgaard considers and rejects two possible answers. The first is to appeal to our human nature, and assert that our psychology compels us. All humans try to satisfy their desires, so there is no need to ask why we ought to. Korsgaard immediately rejects this kind of answer. She says that such an answer "does not have the

structure of reason-giving" (164). It would be trying to derive an ought, "that we should try to satisfy our desires" from an is, "that we do in fact try to satisfy our desires." It would not *justify* our act of choosing the salad.

Another answer that Korsgaard immediately raises and rejects is that our decision of trying to have the things we desire can be the result of a random choice. This is a strange suggestion, but the idea is this. There are various principles of action that we can choose to adopt. We just randomly adopt one, and we happen to choose this principle of making it our end to satisfy our desires. And so we might not need any reason at all to choose this principle of trying to satisfy our desires. But according to Korsgaard, choosing a principle by which to act is itself an intentional action; and so, it too requires a reason. Randomly choosing a practical principle does not, according to Korsgaard, constitute a reason for making that choice. She writes that making such a choice would be "inconsistent with the very idea of a will, which does what it does according to a law, or for a reason. It seems as if the will must choose its principle for a reason and so always on the basis of some more ultimate problem" (164). Her response here, I think, needs more development, because sometimes randomly choosing what to do is part of a reasonable decision making process. For instance, two chess players can agree that whoever wins the coin toss gets to choose who goes first. There is nothing odd or irrational about accepting the coin toss's results; in fact, it seems that accepting the results is the rational and obligatory thing to do, provided the coin toss was fair. But I suspect that there is a good answer to this worry, perhaps, because in the chess player example, there is still a further reason that justifies the particular randomness, and that reason does not depend on randomness. I believe that Korsgaard thinks that our *fundamental* principle of action cannot be the result of random decision, and I think that's plausible.

One possible route of trying to find an answer to the question "why satisfy our desires" that Korsgaard does not consider is to say that the principle in question is self-justifying.

By "self-justifying", I do not here mean "self-evident", but rather that it justifies itself. Consider the principle in question: I will make it my end to have the things I desire. This principle has a self-justifying quality—it helps me satisfy my desires if I adopt the principle of trying to satisfy my desires. This isn't exactly circular, because it's not the repetition of a premise, but the application of a rule to itself. We can see that it is more than just circular by considering all alternative candidate principles, which couldn't be self-justifying in the same way. For instance, consider the following principle: I will make it my end to have *nothing* I desire. Though adopting such a principle will help me leave my desires unsatisfied, in so doing, it gives rise to a contradiction. Making it my end to have nothing I desire becomes one of my desires, and it is not one I can aim to satisfy without contradiction. Once I aim to satisfy it, I violate the principle that I make it my end not satisfy any of my desires. This principle turns out be self-defeating. The original principle in question—that I will make it my end to have the things I desire—is not so self-defeating. I'm not sure how far this line of thought can take us, but I suspect it to be limited. One serious difficulty with this proposal is that there are other similarly self-justifying principles. For instance, "I will make it my end to have many things I desire so long as my desire is not unreasonable" is similarly self-justifying. Perhaps "self-justifying" is too strong; a better word may be "self-supporting" (or at the very least, "not self-defeating"). The word "justification" suggests that a good reason has been given to believe something to be right, and given the possibility of alternative principles equally self-supporting, the self-supporting nature of the principle "I will make it my end to have the things I desire" doesn't suffice as justification. We would need to show that it has better grounding than the others.

Korsgaard abandons the attempt, and tries to assert a different practical principle instead. She says that the only possible way out of this regress argument is to be able to provide a practical principle (or a justification of a practical principle), of which it would be non-sensical to ask why should one adopt that practical principle (or non-sensical to

ask why should one accept the justification of that principle). Korsgaard thus offers the following practical principle which is necessary for any intentional action: "All that it has to be is a law" (166). The idea behind her suggestion is that all intentional action is action for a reason. As we have seen, reasons, in order to count as reasons at all, must incorporate some abstract principle. The abstract principle that Korsgaard has in mind, the one that underwrites all intentional action, is simply that one must act in accordance with an abstract principle, regardless of what it may be. Imagine, Korsgaard says, a rational being trying to adopt a rational principle of action prior to having chosen any practical principle at all. The only constraint on such a rational agent is that it must be a practical principle. Since there is no content for such a rational being, the rational being's only practical constraint is that it follow some principle. Or in her words, "all that it has to be is a law." This would apparently circumvent the regress problem. Since intentional action requires a reason, and a reason requires a formal principle, to ask why you should adopt such a principle (that all it has to be is a law) as your own practical principle is tantamount to asking for a reason to accept reasons as guiding. This question, according to Korsgaard, is non-sensical.

And according to Korsgaard, her proposed principle—all that it has to be is a law—is just the Formula of Universal Law (FUL), Kant's first formulation of the Categorical Imperative. FUL tells us to take our maxims and make sure that we could will them as universal laws of nature, but apart from that constraint, it doesn't tell us what kinds of maxims are acceptable. I think that this is an attractive account of Kant's move from rational agency, through negative freedom, to the moral law. And I think it is likely this is what Kant means when he asserts that negatively free, autonomous, rational beings with wills take the moral law as their own. But it suffers from the problem that Hill pointed out. So Korsgaard's conclusion is that the only rational constraint on intentional action is that it has to take the form of a law. But it isn't clear, at least not to me, that "all that it has to be is a law' is equivalent to FUL.

FUL is a moral principle which issues substantive moral results. It asks us to consider the maxim that underlies our proposed action, imagine a world where that maxim is a universal law of nature such that everybody always acts in its accordance, and to consider whether we could rationally will acting on our maxim in such a world. On the other hand, Korsgaard's principle that "all it has to be is a law" doesn't seem to imply all that; rather it only suggests that we must be able to regard the principle we act on as a law. All it means is that it be a reason, that it can be made abstract, without logical inconsistency. That is, whatever maxim you adopt, you can generalize on the particulars, such that it can be held true. Suppose my reason for speeding is that I am in an emergency. My maxim might be this: whenever I am in an emergency, I may speed. The law would be "whenever anyone is in an emergency, she may speed." But the claim that "all it has to be is a law" is much weaker than FUL; the former allows a far greater range of maxims than FUL does. For instance, we can see the disparity between the two by considering Kant's own example in the *Groundwork*, in particular his fourth one, which is an attempt to show how FUL can issue positive and imperfect duties that we have to others. Kant considers the situation of a man "for whom things are going well while he sees that others (whom he could very well help) have to contend with great hardships, thinks: what is it to me?" (G 423) The man considers adopting the following maxim: I am not to do anything in order to help others when they are in need. Let us call this the maxim of indifference, following Rawls (174). There seems to be no inconsistency if this man would be willing to accept that others treat him the same way. And further, as Kant himself admits, a world where such the maxim of indifference "became a universal law the human race could admittedly very well subsist" (G 423). But he also holds that it is "impossible to *will* that such a principle hold everywhere as a law of nature" (G 423, emphasis in original). There is no consensus on how to understand Kant's reason for thinking that FUL prohibited adopting such a maxim. One common suggestion is that if it were a universal law of nature, it would conflict with our

imperfect duty to cultivate our own happiness (Rawls 174, Sullivan 53). But there seems to be no logical contradiction in the maxim of indifference being a universal law, and Kant even grants that the human race might survive well enough even if it were a law of nature. So nothing prevents it from having the form of law, or being a law, because at this point in Korsgaard's argument, acting in accordance to a law is just the same as acting for a reason. There seems to be nothing essentially irrational about adopting the maxim of indifference.

There is thus a gap between being able to act for reasons and being required to act according to the Formula of Universal Law. Korsgaard tries to close the gap by suggesting that acting according to FUL is constitutive of acting for a reason. She writes that FUL "merely tells us to choose a law. It's only constraint on our choice is that it have the form of a law. Nothing provides any content for that law" (166). I think there is something appealing about thinking of FUL in such a way, and Kant seems to say things that suggest such an interpretation. But it's hard to see how it can be true.[16]

2.5.2 Siding with the Noumena

Korsgaard offers an interesting interpretation of the Intellectual World Argument. According to Korsgaard, Kant's Analytic Argument doesn't quite establish our obligation to the moral law. We are, after all, only human beings; we are mortal and finite animals. The conclusion of the Analytic Argument is that rational agents, not human beings, take the moral law as their own law.

Korsgaard believes that Kant's Analytic Argument succeeds in establishing that the moral law is something we are capable of following, but we need a further argument to show

[16]The argument here in the Analytic Argument is very similar to the argument Kant offers at the end of his first chapter. And at the end of *Groundwork* 1, Kant writes, "Since I have deprived the will of every impulse that could arise for it from obeying some law, nothing is left but the conformity of actions as such with universal law, which alone is to serve the will as its principle." He then articulates FUL (G 402). Even though I am certain that Kant takes himself to be offering something new in *Groundwork* 3, it is unclear to me what that is.

why we human beings would feel compelled to do so. The Analytic Argument tells us that rational beings are capable of following the law. (Or maybe it tells us that rational beings recognize the truth of the claim that they ought to follow the Categorical Imperative.) Kant is here supposedly trying to explain our interest in the moral law, which he believes has thus far been unaccounted for. But phrasing his worry is a little tougher than it first seems. By "interest" does he mean motivation or obligation? Is he asking for a psychological account of what could motivate us to act morally? Or is he asking for an account of what makes the moral law obligatory? Kant is not clear. He writes,

> [F]or, if someone asked us why the universal validity of our maxim as a law must be the limiting condition of our actions, and on what we base our worth we assign to this way of acting—a worth so great that there can be no higher interest anywhere—and asked us how it happens that a human being believes that only through this does he feel his personal worth, in comparison with which that of an agreeable or disagreeable condition is to be held as nothing, we could give him no satisfactory answer. (449-450)

The question Kant's interlocutor seems to be posing is this: Why should we act according to the moral law? Why do we think acting according to the moral law has such high worth to begin with? Let us exclude one answer immediately. Kant's interlocutor is not asking why it is prudential or in our benefit for us to act according to the moral law. This interlocutor is taking Kant's ethics seriously. He is not asking what Glaucon asked—why is it in our benefit to act morally? But there is still an ambiguity. The interlocutor could be asking for an explanation of our interest in the moral law or he could be asking for a justification for the moral law. (One might think that he is clearly asking for a justification, because he phrases it normatively: "why the universal validity of our maxim as a law must be the limiting condition of our actions." But we can ask "why must the planets move around the sun according to Kepler's laws?" and we can ask "why should our bodies deteriorate without food?", without intending that there is anything that the planets must do or our bodies should do.)

Regardless of the ambiguity, Kant seems genuinely concerned with the interlocutor's concern, enough to think that a circle looms. Kant seems to believe that the only answer is that we can think that the moral law has value only if our freedom presupposes a moral interest. But the Analytic Argument started with freedom and then went to the possibility of acting in accordance with the moral law. And hence, we have not yet accounted for our moral interest. Kant's solution invokes the distinction between appearances and things in themselves. But it is far from clear how it is a solution to anything.

Korsgaard offers an interpretation, but her interpretation doesn't really settle my question: Is he offering an explanation of our interest in the moral law or a justification of it? She does not explicitly state what she intends nor can I make it out from her reconstruction of Kant's argument. If I understand her correctly, her reconstructed argument goes as follows. First, the phenomenal world is completely (empirically) causally determined. Each event in it is determined by prior events and the laws that govern their relation. Each of our inclinations or desires is a mental state, which falls within the phenomenal world. Choosing to act to satisfy a desire or inclination is to align ourselves within the phenomenal world. But we have another option. Since the moral law is a demand that asks us to act independently of the desires or inclinations we happen to have, we can choose to side against the phenomenal world—making ourselves more than just phenomenal objects. Since Kant holds that the noumenal world is the 'ground' of the phenomenal, by acting morally, we choose to align ourselves with the noumenal world. We become part of the ground of the phenomenal world. In making decisions, we are now faced with the option of either being at the whims of the phenomenal world or being something that helps make the phenomenal world what it is.

I still cannot tell whether this is supposed to explain why we ought to align ourselves with the noumenal world or whether it motivates us to align ourselves with the noumenal world. But either way, it doesn't seem to work. Korsgaard seems to be making an offer:

either we become the playthings of the universe or we stand up and be something real. Be part of the phenomenal world or be part of the noumenal world. But like all offers, it only seems tempting if we already have an interest in what is being offered. Yet that interest was the very thing to be explained. We have gotten nowhere. Either obey my commands or rebel. But unless you already want (desire or believe that you ought) to be self-legislating, obeying looks just as good as rebelling.

I wish I could offer an interpretation of Kant's argument that does make sense, but I confess that I do not understand it.

2.6 Conclusion

Here's what we've learned. The Analytic Argument is an argument that attempts to establish our obligation under the moral law from our negative freedom. By "negative freedom", Kant does not mean freedom in the sense that it can conflict with determinism. To be negatively free means to be able to act for reasons which are independent from the desires and inclinations we happen have. This conflicts with a thorough-going belief-desire psychology (that beliefs and desires are sufficient conditions for intentional action), but not with determinism. The claim of negative freedom is thus a claim about practical reason, and not a metaphysical claim about our place in the causal world.

Second, I noted that autonomy, for Kant, has at least two components: (1) that an autonomous being can act according to a principle which is not based on any desire or inclination that the agent happens to have, and (2) that the very principle by which an agent acts comes from the agent herself. Aspect (1) of autonomy follows from negative freedom if we assumed that acting for a reason requires acting in accordance with some generalized principle. Aspect (2) of autonomy, I suggested, is ambiguous between a normative/practical reason claim and a metaethical claim. Read as a metaethical claim, (2) would require that

what makes a principle true or valid must come from the agent herself. But this would not follow from assuming negative freedom and the first aspect of autonomy. Rational intuitionists readily claim that a person can act for reasons that do not depend on the desires and inclinations she happens to have, but those reasons will depend on external features of the world, rather than internal features of the world. There is no obvious contradiction in that.

By (2), I think Kant intends to claim that the acceptance of such a principle cannot be motivated from external sources. It is a psychological claim about what can motivate us to action. We can in fact recognize the authority of certain demands without assuming that the authority rests on any particular desire or inclination. But this leaves open what makes the demand valid. What makes a moral claim upon us true is separate from what motivates us to act in accordance with it. And if Kant is right, that the motivation to act in accordance with a moral claim cannot depend on desire and inclination, then the motivation for following it must somehow come from the agent herself. This may seem mysterious, but given that Kant's supposition of negative freedom already precludes a belief-desire psychology, its mystery is not that it cannot be done, but only that we don't know what it is. In various places in the second *Critique* and in a footnote in the the *Groundwork* (401), Kant believes that the moral law can stir within us a feeling of respect. At any rate, I think it is best to read Kant's argument for autonomy as not implying that what makes something right or wrong is us, or at least it shouldn't be implying that yet.

Another lesson from trying to understand Kant's Analytic Argument is that Kant doesn't suppose that there is such a thing as an instrumental principle or a Hypothetical Imperative, as something we ought to follow. There are hypothetical imperatives, and they do take on a general form. But being required to take the necessary means to what we will is constitutive of the concept of willing. Kant does not hold that we are rationally required to subscribe to an abstract principle, such as take the necessary means to one's chosen ends. There is

no instrumental principle that we are obliged to follow. We are under certain hypothetical imperatives because we will things. If we didn't will anything, we would not be under any hypothetical imperative at all. If there was a principle of instrumental reason which we ought to accept, the step from negative freedom to our autonomy would be nigh impossible.

Another point I tried to make was regarding the Under the Idea of Freedom argument. When we read "freedom", not in Kant's peculiar concept of negative freedom, but in the more usual metaphysical sense, we are faced with showing that we can simultaneously think that our actions and thoughts are both entirely determined, from a theoretical perspective and not so determined, for the purposes of deliberation. I argued that it would simply be easier to read the Under the Idea of Freedom argument as an argument about negative freedom (which is a capacity described in terms of practical rationality), and not about free will (which may suppose that we can act differently from what we do). That way there is no reason to think that negative freedom and determinism conflict in the first place. But even still, reading the Under the Idea of Freedom argument as setting a qualification, "from a practical perspective", on some of Kant's conclusions makes it harder to derive those conclusions, even if so qualified they become more palatable.

I think a general lesson regarding how to read Kant's *Groundwork III* can be drawn. It is better to read it as much as possible as not making metaphysical claims, either about the nature of morality or about free will. For instance, we shouldn't expect that the concept of rational agency contains within it the falsity of rational intuitionism. And I don't think that Kant's analysis says so.

So on the whole, I think that Korsgaard correctly interprets the step from the autonomy to the moral law within Kant's Analytic Argument. I think Hill is right to interpret negative freedom as a capacity to regard one's justifying reasons for action to not always depend on the desires and inclinations one has.

However, I do not think that Kant's Analytic Argument works. We cannot bridge the

gap between being rational and standing under the moral law (understood as the Categorical Imperative). My analysis of Korsgaard's interpretation shows where the gap remains. I do not think that the Under the Idea of Freedom argument works, and it is unclear to me exactly what Kant's argument is meant to be, if he even has one. I also expressed my doubts about the Intellectual World Argument. Korsgaard's interpretation is interesting, but it fails as an answer to the question of our interest in the moral law. Whether it is a correct interpretation of Kant, I am undecided.

The remaining questions for Kant, as far as I see it, are these: How do we know we are negatively free? How do we recognize the Categorical Imperative as morally obligatory? And how are we motivated to follow it? Kant's answers to these questions in the *Groundwork* are, as I have tried to show, unsatisfactory.

Chapter 3

The Fact of Reason in the *Critique of Practical Reason*

3.1 Introduction

In the *Critique of Practical Reason*, Kant tries to articulate a conception of morality that is free of the problems found in the third chapter of his own earlier *Groundwork*. It is however generally accepted that Kant, in his later works, continued to endorse the findings he made in the first two chapters of the *Groundwork*. So, naturally, contemporary Kantian ethical theory rests much more on the first two chapters of the *Groundwork* than on the third and final chapter. But most of this theory is normative, and very little of it is concerned with its justification.

As you'll recall from my earlier discussion, chapters 1 and 2 of the *Groundwork* attempt to specify the content of the fundamental moral law, on the assumption that it exists. Only in the third and last chapter of the *Groundwork* is Kant concerned with trying to show that the fundamental moral law is real and valid for us human beings. I argued that Kant's argument there fails, and I gave an account of what its failure consisted in. Most scholars seem to agree that it does fail, even if they have different accounts of its failure. Moreover, most scholars take Kant's argument in the *Critique of Practical Reason* to provide an

improvement, but there seems to be even less agreement on what Kant's argument there is.

The basic outline of the argument in the third chapter of the *Groundwork* begins with the fact of our rationality, and then argues for our freedom, and finally, for our subjection to the moral law. In the second *Critique*, however, Kant seems to reverse the argument. He doesn't explicitly mention the *Groundwork* or its argument, but he says things that suggest that it was bound to fail. In particular, Kant denies the possibility of proving our freedom via theoretical means.

Paul Guyer, in fact, seems to think that this switch was inevitable, because the first *Critique*, published a few years before the *Groundwork*, already denies the possibility of success for the kind of argument the *Groundwork* presents. Writing about the *Groundwork*, Guyer says that Kant "simply seems to assume what the first *Critique* had denied.... Kant offers no justification for this sudden departure from the epistemological constraint that is central to the argument of the first *Critique*" (222). This interpretive explanation—that Kant would so obviously contradict himself—is hard to accept, but the idea that Kant became disappointed with his treatment of freedom and our knowledge of the moral law in the *Groundwork* is not.

In the first *Critique*, Kant argues that we cannot have knowledge of our freedom. According to Kant, the world, as it can be known empirically, is completely determined, and it has no room for the kind of freedom morality requires. So, Kant's reversal of argument, in the second *Critique* begins instead with the fact that we are subject to the moral law, and then he shows us that we are free. But this reversal raises many questions. How and why does being subject to the moral law imply that we are free? Can this freedom simply be a compatibilist conception rather than the transcendental freedom that Kant seems to be arguing for? And most importantly, how do we know we are subject to the moral law?

Moreover, this move may seem disappointing: what was originally a difficult and rich, though flawed, argument for the claim that we are subject to the moral law, now turns

on what appears to be mere assertion. The very claim he was so desperate to prove is apparently simply granted. It is unsurprising then that various philosophers have offered various reconstructions and interpretations of Kant's argument that render it to be more than mere assertion. Most prominent among them are Lewis White Beck, John Rawls, and Henry Allison.

In each of these interpretations, Kant's argument depends on something he calls "the fact of reason". But it is far from clear what this fact is. Kant is neither always perspicuous nor even consistent. The three most significant passages concerning what the fact of reason is seem to be these:

> (1) Consciousness of this fundamental law may be called a fact of reason because one cannot reason it out from antecedent data of reason, for example, from consciousness of freedom (since this is not antecedently given to us) and because it instead forces itself upon us of itself as a synthetic a priori proposition that is not based on any intuition, either pure or empirical .... (5: 31)
>
> (2) This Analytic shows that pure reason can be practical – that is, can of itself, independently of anything empirical, determine the will – and it does so by a fact in which pure reason in us proves itself actually practical, namely autonomy in the principle of morality by which reason determines the will to deeds. At the same time it shows that this fact is inseparably connected with, and indeed identical with, consciousness of freedom of the will... (5: 42)
>
> (3) Moreover the moral law is given, as it were, as a fact of pure reason of which we are a priori conscious and which is apodictically certain, though it be granted that no example of exact observance of it can be found in moral experience. (5: 47)

It's not obvious how these various formulations of the fact of reason are equivalent. In the first passage, Kant says that the fact of reason is "consciousness" of the fundamental law of morality. In the second, the fact of reason is "autonomy in the principle of morality" *and* "consciousness of freedom of the will." And in the third passage, it is "the moral law" itself. One might think that there are many different 'facts' of reason, except that Kant also says that it is the "*sole* fact of pure reason" (5: 31). So there is first a question about

interpretation: what exactly is the fact of reason? Also, the arguments that start with the fact of reason seem to be deep and interesting, and they suggest a change from the argument in the *Groundwork*. But what is that argument? In this chapter, I will examine various prominent proposals for answers to these questions. But let us see if we can glean what is certain.

Kant seems to be saying, in these passages, that there is some fact of which we are aware, but somehow our awareness of this fact is not given to us via the senses or even by intellectual intuition. This fact is supposedly synthetic a priori, and so we are conscious of it and certain of it. Yet Kant maintains that we have no observance of it in our moral experience. This ought not to seem possible. Kant says that we have no intuition of it nor can we prove it using theoretical means, but we know it anyway. He says that the moral law "provides a fact absolutely inexplicable from any data of the sensible world and from the whole compass of our theoretical use of reason" (5: 43).

Moreover, whatever fact this is, it is able to show to us our freedom—and not just a compatibilist conception of freedom, but a transcendental one—a noumenal one that is independent of the empirical world.

Upon first reading, one might have the following interpretation: We face decisions that we have to make, ones between decisions which serve our self-interest and those which ask us to make our wills subservient to the moral law. That is, sometimes there is a conflict between what we *want* to do and what we *ought* to do. We see in ourselves the capacity to choose to act in accordance with what we ought to do, instead of what we want to do. This means that we can act in accordance with the moral law. Moreover, since the moral law has its authority independent of the desires and inclinations we happen to have, we do not have to act in accordance with our desires. We are thus transcendentally free.

But this interpretation suffers from various difficulties. First, it says that we see in ourselves the capacity to choose in accordance with what we ought to do. But this could be

an illusion, so the sceptic might argue and as Kant allows. There are two possible sources of scepticism. The first scepticism consists in denying that any of us can see in ourselves or in others the capacity to act as we ought to, but rather, if we look closely enough, we will only find a capacity to act in our own self-interest. Such a sceptic professes a psychological egoism, and claims potential counterexamples as mere delusions. You may think that your motives are pure, so says this sceptic, but they are not—believing that you act for reasons that are noble and not ultimately self-interested is a kind of bad faith. A second source of scepticism may begin with hard determinism. A hard determinist may say that we do indeed have this impression, but that it represents no genuine capacity. It may appear to us that we can act as we ought to, even when we do not, but that is the illusion of freedom. If you did as you ought, so says the hard determinist, then you did the only thing you could have done. Likewise, if you did as you ought not, then you did the only thing you could have done. There is no genuine capacity to have done otherwise. It is not because you lack that capacity (whereas others might have it), but the universe and its laws, as they are, do not allow for it. Similarly, we cannot travel faster than the speed of light, not because we human beings, in virtue of being the kinds of creatures we are, lack some capacity, but because the laws of the universe do not allow for it. And so, our impressions of free will must be illusions.

There is a second problem with the interpretation sketched above. Even if we can resolve these sceptical doubts about such a capacity, there is another. If we were certain that we can choose to act in accordance with duty rather than self-interest, it does not follow that we need anything resembling transcendental freedom to account for that possibility. A classical compatibilist conception of freedom may be all that is needed, which may hold that our choices to act morally are compelled by inner mechanisms that do not always serve one's self-interest. Not all our desires, according to the compatibilist, need to be self-directed. According to the classical compatibilist, we could have acted in accordance with

the moral law even when we didn't, and thus we can be held morally responsible. Classical compatibilists assert that our actions are free when we perform them from our desires, motives and character traits. Our actions are unfree to the extent that we are constrained by physical forces, coercion or inabilities. Furthermore, just because the moral law's authority does not depend on the desires and inclinations we have happen to have, it does not follow that we cannot act out of a desire to act morally. So according to the classical compatibilist, we can perform a moral act with the right motive, because we have some desire to act morally. And in particular, for the classical compatibilist, even when we do something we ought not to have done, we could have acted otherwise. But this capacity to act otherwise does not require transcendental freedom, because for a classical compatibilist, we always could have acted otherwise if we were free from external constraints to do so.

Third, this interpretation seems to work even if the moral law were something other than what Kant thought it was—if, say, it was the principle of utility. There seems to be nothing special about the Categorical Imperative that is essential to this argument. If I believe that the principle of utility is the one true moral law, I can recognize the fact that what it demands of me may conflict with what serves my self-interest. And it appears that I can choose to act in accordance with this principle, instead of, say, the formula of universal law. If so, then my recognition of the principle of utility as authoritative for me can imply, in just the same way, that I am free.

And so, in order for Kant's argument to succeed, it requires better interpretation.

3.2 Beck's interpretation

Lewis White Beck believes that there are two arguments, rather than one, that lead Kant to justify our belief in morality, and both depend on the fact of reason. Beck's first argument involves explaining what the fact of reason is and what it justifies, while the second involves

relating the fact of reason to Kant's views about the nature of theoretical reason.

3.2.1 Beck's first argument

First, a bit of setup. Kant believes that people, in general, have a certain respect for moral considerations. Kant believes that the starting point for investigating the nature of morality is our ordinary moral consciousness—in particular, our common sense views about the nature of morality. But Kant admits the possibility that our ordinary beliefs about morality may not actually refer to anything—that morality does not, in fact, exist, and that our ordinary beliefs about right and wrong, our belief in our own free will constitute only a "vain delusion" (G 402). Our ordinary moral beliefs tell us what morality must look like, if it existed, but we must still prove its reality.

If we look at the various passages above, we can distinguish among those that describe the fact of reason as *consciousness of* X and those that describe it as X itself, where X is the moral law, or autonomy, or freedom. Kant's term "fact of reason" or "fact of pure reason" appears to be ambiguous between two very different things—our awareness or consciousness of a particular thing, and the thing itself. Perhaps what needs explaining is why Kant allows himself this equivocation. One might easily admit that we are 'conscious' of the moral law, understood in the qualified sense that it appears to us that moral considerations factor into our lives, without admitting the stronger metaphysical claim that the moral law actually exists or is actually authoritative for us. Likewise, one might easily admit that we experience the phenomena we associate with free will, but not thereby admit that we have free will. What we need is to explain Kant's move from one to the other.

Beck proposes the following: Kant's term "the fact of reason" or "the fact of pure reason" can be distinguished between a fact *for* pure reason and a fact *of* pure reason. The fact for pure reason is so-called because it would appear to be a fact given to and

apprehended by pure reason. It is almost something we are directly aware of. And we might even be tempted to call it an intuition, in Kant's sense of the word, as if it were part of the given. The apparent fact for pure reason, the thing we are almost directly aware of, is that we appeal to moral considerations in our practical decision-making.

The fact *of* pure reason, on the other hand, is "the fact that there is pure reason known by reason reflexively" (Beck 168). Beck's argument is then this:

> Only a law which is given by reason itself to reason itself could be known a priori by pure reason and be a fact for pure reason. The moral law expresses nothing other than the autonomy of reason; it is a fact for pure reason only inasmuch as it is the expression of the fact of pure reason, i.e., of the fact that pure reason can be practical. That is why the moral law is the fact of pure reason and for pure reason. (169)

This is a strange and terse argument, and it is unclear that it works or how it is meant to work.

Roughly, we start with the claim that if a law can be both *a priori* and a fact of pure reason, then it must be given by reason itself. Then we note that we have this particular moral phenomenon where it appears to us that we can choose to act in accordance with what we take to be morally right course of action.

Beck thus concludes that this moral phenomenon is an expression of pure reason being practical. Or rather, if pure reason were practical, it would manifest itself as this kind of moral phenomenon. But this argument as it stands cannot establish that pure reason is practical. It reasons from phenomena to an explanation, where other explanations remain possible. In short, it appears to be an inference to the best explanation.

But in order for an inference to the best explanation to count as a good argument, we need to show that of the possible explanations, it is the best one. That is, it has to meet particular standards better than other candidate explanations. The usual standards for what is to count as best include simplicity, elegance, and explanatory power, among others. And

even though this leaves some wiggle room for what is to count as best, it is not obvious that pure reason being practical is the best explanation of the moral phenomena in question. Another possible explanation of the phenomena of taking moral considerations is the compatibilist one, which says that humans have a desire to act morally, and since our desires figure in our practical deliberation, this phenomenon is hardly peculiar. And since Kant has no reason to object to the claims that we have a desire to act morally or that our desires figure in our practical deliberation, the proposed Kantian explanation here—that pure reason is practical—is hardly simpler. It requires more assertions to accept. (As a mere matter of what is to count as best explanation of phenomena, whether Kant is right to think that acting out of desire to act morally has moral worth or not is likely to be irrelevant.) Or another explanation of the phenomena is the hard determinist's one, that this phenomenon—that we sometimes make moral considerations in our practical deliberation—is only appearance. We do not actually take moral considerations seriously in our practical deliberation, it only seems as if we do. A trick of the brain, so the hard determinist might say, can cause this seeming. The bare fact *for* pure reason and the fact *of* pure reason thus seem to connect tenuously.

Beck, however, does not describe his argument as an inference to the best explanation, and he seems to want to say something stronger, something approaching strict validity. One way for this argument to work would be to specify a particular characteristic of moral phenomena that other explanations cannot account for—perhaps the required characteristic would be the recognition of the Formula of Universal Law (or some other formulation of the Categorical Imperative) as the fundamental principle of morality in our moral consciousness. This, however, would require showing that the recognition of FUL as the fundamental principle of morality as somehow equivalent to pure reason being practical. And it would require assuming that people do, as a matter of fact, recognize FUL (or some other formulation) as the fundamental principle of morality. But many people (or perhaps even most

people), as a matter of fact, do not actually take FUL to be the fundamental principle of morality.

Fortunately, Beck offers a different elaboration of his argument, one that aspires to deduction rather than abduction. But unfortunately, the argument is terse and unconvincing. Beck writes,

> A moral principle is not binding upon a person who is ignorant of the principle or law. On the other hand, if a person believes that an imperative is valid for him, then it is in so far forth valid for him, and he shows that reason is practical even in the awareness of this respect of a valid claim. (169)

There seems to be something to this. But the way Beck writes it seems false. The word "binding" is ambiguous between a motivational aspect and a normative or obligatory aspect. Under the normative, and I think more natural, reading of "binding", it is false that moral principles bind only those who are aware of them. Laws are binding on people even if they are ignorant of them. People ought not to drive drunk, even if they do not happen to know that it is wrong. People ought not to drive on the left side of the road, even if they do not happen to know that it is against the law. People are bound by laws, morals, and even etiquette, regardless of whether they are aware of them or not.

But nevertheless, what Beck says seems to have some truth in it. There seems to be something to his claim that "if a person believes that an imperative is valid for him, then it is in so far forth valid for him". But it is not easy to make this clear. Part of it is motivational, but I don't think that tells the whole story. I want to distinguish between two different threads in Beck's thought. The first is motivational. If a person does not believe that she ought to ϕ, she is unlikely to feel obliged to ϕ. We cannot, however, baldly assert, without argument or evidence, the stronger claim that the belief that one is obliged to ϕ amounts to the feeling, or necessitates the feeling, that one is obliged to ϕ, for otherwise motivational internalism would be true by stipulation.

Beck might think that it is hard to imagine how someone could feel obliged to ϕ without believing, on some level, that she ought to ϕ. It's unclear exactly what Beck has in mind, but what I've specified here won't work for his argument. Since having the belief that one ought to ϕ does not necessitate one *feeling* obliged to ϕ, without argument or evidence, it cannot show that "reason is practical". That is, even if reason alone, without appealing to our desires, could produce an actual obligation, it wouldn't be practical unless it can be shown to move us to act in its accordance. Some non-'reason' element may be needed before it becomes practical, as a Humean might contend.

The second thread is this. It is paradoxical and seemingly contradictory for anyone to say "I believe that I ought to ϕ, but I ought not to ϕ." If this claim is not just *seemingly* contradictory, but *actually* contradictory, then the following claim should have the ring of an analytic truth: "If I believe that I ought to ϕ, then I ought to ϕ."

Perhaps it has the ring of truth, but it is very difficult to explain why it does so. After all, if we change it from a first-person report to a third person report, it turns out to be false. Consider the third-person version:

(α) If S believes that he ought to ϕ, then he ought to ϕ.

This claim is false. The following argument shows why. Suppose that John believes that he ought to kill innocent people for fun. Let us assume, for reductio, that (α) is true. It would then follow, by a simple *modus ponens*, that John ought to kill innocent people. But this conclusion is simply false. It is simply not the case that John ought to kill innocent people for fun. We must then reject the assumption that (α) is true.

And so, I do not see how the first person version is true either. I do not think it follows from the fact that I believe that I ought to ϕ that I ought to ϕ, no matter how paradoxical and contradictory it may seem to deny it.

Beck is fully aware that believing that one ought to ϕ does not make it the case that one

actually ought to ϕ. So what is Beck's point? He writes,

> This is true whether the imperative expresses a claim that is in fact valid or not. Only a being with an a priori concept of normativity would even make a mistake about this. To argue against it is to appeal to normative grounds and is as ridiculous as to attempt to prove by reason that there is no reason. (169)

As far as I understand Beck, this is his point. We take moral considerations as relevant to our decision making. Only a "being with an a priori concept of normativity" (I am understanding this to mean "a being for whom pure reason is practical") could believe that such considerations ought to guide his behaviour—that believing that he ought to ϕ gives him a reason to ϕ.

But if this is Beck's argument, we're in no better a position than we were when we started. It is not obvious to me that only a being with an a priori concept of normativity could make this mistake. Beck has specified the moral phenomena further, so as to include the fact that everyone finds it odd and paradoxical to say "I believe that I ought to ϕ, but I ought not to ϕ". But there is little reason to think that his explanation—that beings for whom pure reason is practical have this belief—is the only available one.

The phenomenon in question is not particular to morality, or even to practical reason. It occurs with belief. Moore once famously pointed out that it is odd and perhaps contradictory for one to assert that "It's raining outside but I don't believe that it is" (1942, 543). Moore's sentence has a similar structure to its normative counterpart. Like in our normative example, there is no contradiction between the claims "It's raining outside" and "I don't believe it's raining." It is possible that both are true, yet one cannot coherently assert or believe both at the same time. Also, there is no contradiction, or any oddity at all, in third person ascriptions: "It's raining outside, but John doesn't believe it."

I suspect whatever explains the oddity of Moore's sentence will also explain the oddity of our normative counterpart. Let us rewrite Moore's example as follows: "I believe that it

is raining outside, but it isn't raining outside."[1] This rewrite is actually more specific than Moore's, because Moore's sentence is ambiguous. The scope of his "not" is unclear. It could either mean "It's raining outside but I do not-[believe it's raining]" or "It's raining outside but I believe that it's not-raining." In either case, a paradox remains, and a solution to Moore's paradox presumably resolves the paradox in both cases.

The normative case remains the same "I believe that I ought to ϕ, but I ought not to ϕ." But both paradoxes now take the same form: "I believe that S, but not-S", where S is some proposition. I suspect that most any instantiation of that form will result in a similar paradox, and hence I suspect that there is nothing peculiar about a particular instantiation of S that would help explain its paradoxical nature. And thus, nothing peculiar about normativity or practical reason will be required to explain the phenomenon.

I am, however, uncertain whether my reply works—for two reasons. One, it is possible that S has to exhibit certain qualities for a Moorean paradox to occur. That is, it may simply not be the case that *any* instantiation of the general form would result in a paradox; there may be a whole class of sentences, when substituted in to the general form, that do not yield anything paradoxical at all. Second, I may be ignoring what is special about the paradox in the normative case that is absent in Moore's case and in the more general case. But seeing as I do not know what sentences would fail to yield such paradoxes, or what special paradoxical quality in the normative case I may be ignoring, I believe that the paradox does

[1]One may worry whether this rewrite captures Moore's thought. Consider someone who says to his psychotherapist: "I believe that my son wants to kill me and marry his mother, but it isn't true." We seem to have no trouble understanding what this means. But I think our trouble with Moore's sentence is not with whether we can understand what it means. Rather, our trouble is with whether the person is being incoherent in saying it. I believe that the patient is being incoherent; in fact, I think it's one of the incoherences that the patient himself wishes to get rid of.

Also, we can see that the structure of Moore's own sentence could be subject to the same kinds of examples. Someone might say to his psychotherapist, "I'm human but I don't believe it," or someone might say to her husband "I won the Nobel prize, but I don't believe it."

So I'm not sure if there is a difference between the psychotherapist example and Moore's sentence. But if there is a difference between Moore's sentence and these examples, I suspect it's due to the differences in context and not in the differing logical forms of the sentences.

not imply, contrary to Beck, that only beings for whom pure reason is practical could make this mistake.

Beck sometimes writes as though this argument, from the fact for pure reason to the fact of pure reason, is meant to be sound. But at other times, it is not obvious that he intends it to be so, especially since Beck offers a second argument. (Perhaps he is unsure of the soundness of the first.) But if the first argument fails, as I think it does, a second argument may be needed. And then, we may ask, why make the first argument at all, if it fails? Perhaps the first argument plays a different role. Perhaps, the fact *for* pure reason, as Beck understands it, can only suggest, rather than prove, the fact *of* pure reason, and that more is needed to substantiate it.

The fact of reason, for Beck, consists in relating two distinct ideas, the fact for pure reason and the fact of pure reason. That is, our recognition of moral considerations in our practical deliberation (the fact for pure reason) is likely a manifestation of the fact that pure reason is practical (the fact of pure reason). I tried to show that Beck doesn't get much further than showing that our consciousness of moral considerations *can*, but need not, be explained by the fact that pure reason is practical. But for Beck that may be enough, given his second argument, which may work in concert.

The first argument says that we have moral phenomena which give us a reason, if not a definitive reason, to think that pure reason is practical or that the moral law is real, which is, for Kant, one and the same thing. So we have some warrant to believe in the moral law.

3.2.2 Beck's second argument

The second argument that Beck offers connects the first argument with a result from theoretical reason. According to Beck, the moral law, if real and true, "serves as ground for the deduction of freedom" (173-174). Of course, all of this remains tentative, because we

do not know, with certainty, that the moral law is real and true. But, fortunately, we have something that comes from the opposite direction. Theoretical reason gives us the idea of freedom.

So how does theoretical reason give us the idea of freedom? Let us briefly recall Kant's third antinomy, and his solution to it, from the first *Critique*. An antinomy is the opposition of two claims, a thesis and an antithesis, both of which seem to have equal justification, whose Kantian solution involves first explaining how both have equal rational justification, and second, how they must be re-understood. So the thesis in Kant's third antinomy is that natural causality is not the only kind of causal power, but that there is freedom which is its own causality, separate and distinct from natural causality. (I won't go into Kant's 'proof' of this claim.) And the antithesis is the more familiar view that there is no freedom, and that causality through laws of nature is the only kind of causality. (Again, I won't go into Kant's proof of this claim.) Kant's resolution of this antinomy involves asserting two claims: (1) we cannot know that natural causality is a law outside the world of experience, but it is true for the empirical world; (2) freedom remains possible in things-in-themselves, for we cannot rule that out. In this way, Kant tries to give some merit to both thesis and antithesis, while qualifying both so that they are compatible.

A corollary of his resolution is that science, which studies the empirical world, can never show us the reality of freedom. And, in fact, whenever we do science, we must pre-suppose that freedom does not exist, and that natural causality is the only kind of causality there is. But since we cannot prove, in principle, that the empirical world is all there is, we cannot rule out freedom. Thus, the existence of freedom is possible. But, for Kant, it is more than just possible. Since the resolution of an antinomy means that both thesis and antithesis have partial merit (after all, Kant does not show the proof of the thesis to be false, but only that its scope and its applicability is limited), we have some non-definitive reason to think freedom is real, and likewise, but to a greater extent, we have some reason to think

that the law of natural causality is true. According to Beck, "theoretical reason thinks that freedom is possible not merely because it can discover no valid logical evidence against it; it *requires* us to think of freedom as possible (174, emphasis his). Reason, according to Beck, requires the possibility of freedom, because without it,

> reason's interest in thinking a connected world would have been thwarted. Reasoning demands a totality of conditions, and if a cause which is not also an effect were shown to be impossible, this demand could not be met in either the phenomenal or noumenal world. (174)

But, as Beck recognizes, that does not give us sufficient reason to think freedom is real; we cannot prove its existence via theoretical means, because applying reason to things outside the possibility of experience is overstepping its bounds. Moreover, we do not have intuitions of freedom, and in fact, we cannot. We cannot have intuitions of things-in-themselves, which is where freedom, if it exists, must reside. So we have an "idea" of freedom, but not freedom itself.

Here then is how practical reason and theoretical reason come together. On the side of practical reason, we have some non-definitive claim to the reality of the moral law, given to us by the fact of reason, provided that Beck's first argument is insufficient to establish the reality of the moral law. And according to Kant, the moral law (or our subjection to it) implies that we are free.

So now we have some partial, but non-definitive, claim to our freedom. And then, on the other side, the side of theoretical reason, we have some non-trivial, but non-definitive, claim to the reality of freedom given to us by Kant's resolution to the third antinomy.

We have two shaky supports that come together at the idea of freedom. As I understand Beck's idea, this is how freedom serves as "the keystone of the whole structure of a system of pure reason, even of speculative reason" (CPrR 5: 3-4). According to Beck, this coming together, this cohering, serves as justification, and this coherence is the "credential" for the

moral law. Beck writes

> The fundamental principle [the moral law or the principle of pure practical reason], already asserted as a "fact," is not left a naked and isolated assertion or an assertion surrounded by a closed, circular, and empty system. . . .
>
> Now, because the Idea of freedom was required but not confirmed by theoretical reason, yet is confirmed practically through the fact of pure reason, there is no danger of conflict. Rather, there is mutual support, which far surpasses the evidential value of mere consistency. The independent warrant of the concept of freedom—to wit, that it is needed also by theoretical reason—makes it serve as a systematic credential for the reality of pure practical reason. (175)

So for Beck, it seems that practical reason has some standing and weight on its own, but not enough for it to be securely established. But since practical reason gives something that theoretical reason needs, the two connect, making both more stable. Thus, the final part of the justificatory work for the moral law lies in practical reason's coherence with theoretical reason.

If we are to be sceptics, then, of practical reason, we cannot be sceptics who accept Kant's results from his investigation of theoretical reason, because those results provide no basis for such scepticism. In other words, scepticism (of an external sort) has to step further outside, perhaps outside of accepting Kant's views altogether. But for Kant, that would mean stepping outside reason altogether. And criticism from that standpoint would be self-defeating, since the very idea of asking critical questions invokes reason itself.

At least this is how I understand Beck's views. But as I've portrayed them, they are steeped in more metaphorical language than I would like. I would like to try to make it clearer. So first, we have this coherence between the results of Kant's investigation of theoretical reason and the results of his investigation of practical reason. This kind of coherence provides more evidentiary support for the moral law, and everything else too, than what mere consistency can provide. At best, it is akin to two plausible scientific explanations of two different phenomena, which appeal to the same posited entity. This

happy accident gives some additional weight, though not utterly decisive, to both scientific explanations and to the posited entity.

Thus, in the same way, the two separate and independent arguments that lead to our positing of freedom provide support for both freedom and the theoretical frameworks from which both arguments arise. One advantage of this coherence is that it might help defeat certain kinds of scepticisms. For instance, if you accept Kant's results from the first *Critique*, and in particular, if you accept his resolution to the third antinomy, you would have fewer grounds to be sceptical of the claim that freedom exists. Indeed, you may have to welcome it.

Thus, if this is correct, then in order to be sceptical of Kant's justification of the moral law, or to be sceptical of freedom, you would have to reject Kant's solution to the third antinomy, and perhaps more generally, the results of his entire first *Critique*. If you take the first *Critique* to present an entirely coherent metaphysical and epistemological doctrine (though it is not necessary that you take it to be true), then rejecting Kant's claim that freedom is real would require rejecting aspects of the first *Critique*. Doing that, however, would be formidable. A successful argument would require premises from which you can show that one of Kant's claims is false. But those premises themselves would have to be claims that Kant did not disprove or seriously limit. If this is what one would have to do in order to reject Kant's claim of freedom, then the task is surely formidable. Perhaps it becomes impossible to raise any reasonable scepticism at all. Kant's first *Critique* is unquestionably ambitious, and in it, he seeks to understand and describe the limits of reason itself. Since all our knowledge begins with reason, no claim is outside its scope. Hence, to question the reality of freedom is, in effect, to question the whole of reason, which is an altogether incoherent thing to do.

But if I've characterized it accurately, Beck's interpretation is disappointing. There are various reasons to doubt its success as a justification of the moral law. I will try to classify

two types of sceptics about freedom depending on what they already believe: (1) there are those who accept that the first *Critique* does, in fact, present a coherent metaphysical and epistemological doctrine; and (2), those that do not. And let us distinguish, in the second type of sceptic, two further sorts: (a) those that accept Kant's analysis of the third antinomy (without necessarily accepting the whole of the first *Critique*, or its coherence) and (b) those that do not.

If you fall into category (1), then you may have a harder time being sceptical of Kant's justification of the moral law. But it is far from impossible. The coherence of the results of the first *Critique* and the second *Critique* lies at the intersection of freedom, and rests on little else. In fact, according to Beck's interpretation of the credential argument, it has to. Nothing practical reason offers can, in principle, contradict or confirm what theoretical reason offers, unless, so it seems, they agreed or disagreed on freedom. Fortunately, they agree on freedom. But you can be sceptical of the whole of the second *Critique*. Nothing in the first *Critique* demands accepting the arguments of the second. You can be sceptical of the phenomenon that Beck calls the fact for pure reason, and hold that our common belief that we appeal to moral considerations is misguided or a case of bad faith. You can reject the idea that pure reason can be practical as outright preposterous, and in particular reject that pure reason's practical nature manifests itself in our practical reason, without giving up any bit of the *Critique of Pure Reason.* In other words, the coherence between theoretical reason and practical reason only adds weight if you accept both the results from both sides already.

If you are a sceptic of the sort (2a), and so do not think that the first *Critique* presents a fully coherent doctrine, it is even easier to be sceptical of Kant's justification of the moral law. In addition to the reasons to be sceptical above, you would also be left unconvinced by the threat that the rejection of freedom amounts to the rejection of reason altogether, because that connection between freedom and reason depended on the coherence of the

results of the first *Critique*, which is a claim you reject. And even if you accept that Kant's defense of a limited version of the third antinomy's thesis—that freedom exists and is separate from natural causality—the keystone argument, or metaphor, begins to look less compelling. Rather than seeing it as two supports that come together, it begins to look like two mediocre arguments for the same claim. Adding up mediocre arguments for the reality of freedom does not add weight to that side of the balance. The number of bad arguments for a claim does not make it more likely to be true, no matter how great that number. Even absurd and patently false claims can come with a barrage of incomplete and mediocre arguments.

If you are a sceptic of the sort (2b), you inherit all the possible reasons for scepticism from the previous two, and you have the following further reason to be sceptical. If you do not accept Kant's resolution to the third antinomy, in particular his defense of a limited version of its thesis—that there is a causality other than natural causality—then no coherence is forthcoming. You may even accept other aspects of Kant's resolution. For instance, you may accept that the law of natural causality is confined to our phenomena, or empirical understanding of the world, but need not accept that there is any positive reason to believe in freedom, except that nothing can logically preclude its possibility among noumena. But since mere logical possibility does not indicate need, the coherence being presented is absent.

In the face of all these possible sources of doubt, I am unconvinced that this is the key argument about the fact of reason. But that doesn't mean that this account lacks any value whatsoever. If you happen to be independently compelled by the results of Kant's investigation of theoretical reason and of his investigation of practical reason, then their coherence, their mutual support, is a welcome feature. But if you are not, then their coherence means little.

I believe that Kant has a better argument than this coherence, and I believe that Beck

thought so too. That is why he provided the first of the two arguments. In fact, I think Kant needs a better argument.

3.3 Rawls's interpretation

Rawls's interpretation is sometimes thought to be similar to Beck's second argument (David Sussman 60). But his argument concerning the fact of reason, at least as it is presented in his *Lectures on the History of Moral Philosophy*, is very different. It is tempting to think of any account of the fact of reason as an account of one aspect of Kant's thought, largely independent from the rest. After all, we tend to think that there is general consensus regarding the broad outlines of Kant's ethical theory and the general results of his investigation of theoretical reason, and so we are tempted to think that Rawls's account of the fact of reason is an attempt to solve the same problem that Beck's account is. But Rawls's view is more radical than it might first appear.

There is some general consensus regarding what Kant takes to be morally required and the kinds of reasons that count as good moral reasons, but there is little agreement about how he conceives of the metaphysical and epistemological nature of moral claims. Rawls has a unique and novel interpretation of the metaphysics and epistemology of moral claims under Kant's view, which happens to play an essential role in his account of the fact of reason. Thus I think we cannot fruitfully ignore Rawls's larger interpretation of Kant's moral theory and focus exclusively on what he says about the fact of reason. Rawls's account of the fact of reason, in particular the role it plays in justifying Kant's moral theory, only makes sense, and is only plausible, within his controversial account of Kantian constructivism.

Rawls uses the term Kantian constructivism for two separate theories: his own theory of political justification, which is found in his book *Political Liberalism* and in his paper

"Kantian Constructivism in Moral Theory" (which, despite its title, is more about political theory than moral theory), and Kant's own moral theory. I shall be concerned with the latter.

Kantian constructivism has two parts: its Kantianism and its constructivism. Let us first deal with constructivism. Constructivism in ethics is intended as a metaphysical doctrine, which is meant to contrast with rational intuitionism and moral realism on one side, and with sentimentalism and moral scepticism on the other. Its ambition is to accept that there are no moral truths independent of rational agents, and to argue, nonetheless, that there are objective moral judgments which cannot be reduced to mere expressions of our feelings. It's easy to see the contrast between constructivism, on the one hand, and sentimentalism and moral scepticism, on the other. Sentimentalism holds that our moral beliefs are expressions of our feelings, and moral scepticism holds that our moral beliefs have no objective merit. We can likewise see the contrast between constructivism, on the one hand, and rational intuitionism or moral realism, on the other. Rational intuitionists and moral realists hold that there are moral truths independent of rational agents, whereas moral constructivists do not.[2]

As an example, many contractarian theories in political theory would be considered constructivist. Many of them hold that our political obligations are objective and are not mere expressions of our feeling, but they also hold that we have no (or very few) political obligations or values prior to the social contract. As I mentioned earlier, Rawls takes his own political theory to exhibit Kantian constructivism. Rawls holds, in brief, that a correct set of political obligations, duties, and rights is the outcome of a process of deliberation from a standpoint of what he calls the original position. The process of deliberation is not

[2]Shafer-Landau (2006) defends moral realism, which he takes to be "the view that says that most moral judgments are beliefs, some of which are true, and, when true, are so by virtue of correctly representing the existence of truth-makers for their respective contents. Further, and crucially, true moral judgments are made true in some way other than by virtue of the attitudes taken towards their content by any actual or idealized human agent" (209).

to be understood as a reliable method for discovering what the correct political society is, as if there were an independent truth of the matter; rather the process of deliberation is to be understood as setting our political obligations, duties, and rights. Constructivism, then, presumes that there is no fully determined set of political obligations, rights, or duties prior to the process of deliberation. In Rawls's political theory, there are some political values prior to the construction, and which help justify the constructive procedure. We take the process of deliberation from the original position to be right because we take it to embody certain values we already hold, for instance, equality, fair bargaining, the moral irrelevance of certain contingent features such as skin colour, sex, etc. But by themselves, these values give us little, and the process of deliberation from the original position, which reflects these basic values, gives us the bulk of our political obligations.

Kantian constructivism in moral theory is Kantian because Kant's Categorical Imperative does work similar to that of the original position. Rawls conceives of the Categorical Imperative as applying to us in a constructive procedure, which he calls the CI-procedure. It is a procedure because it involves several steps: first, identifying the maxim that encapsulates your reason for acting as you propose; second, recasting the maxim as a universal law of nature governing all rational agents; and third, considering whether your maxim is conceivable in a world governed by this law of nature and whether you could or would rationally will to act on your maxim in such a world (if not, you have a duty to refrain from acting on such a maxim). (Rawls considers it four steps, but I have paraphrased it into the more usual terms we use to describe Kant's Categorical Imperative, and I have collapsed his steps 2 and 3 into one step (167-169).) The procedure is constructive, because prior to applying it, there are no permissible or impermissible maxims. The procedure makes your maxim acceptable or unacceptable. That is, all there is to whether your maxim is acceptable or not is whether or not it passes the CI-procedure.

In contrast, moral realism holds that there is an independent moral reality, and our

job is to try to figure out what that is. But that task seems to pose certain epistemological difficulties. How are we to learn about this independent reality? Perhaps we have a capacity for a kind of direct moral perception that allows us to have glimpses of what that reality looks like. Or perhaps we can use reason—maybe by doing moral philosophy—to help us figure out what it must be like. Or perhaps there is some other method, or some combination of all the above. In any case, under moral realism, the task of moral inquiry is to figure out what that reality is.

But on the constructivist account, the procedure is not a method of acquiring that knowledge. It does not give us good evidence regarding some independent moral reality. Rather, the procedure produces particular moral truths. Rawls puts the contrast between rational intuitionism and constructivism as follows.

> Rational intuitionism says: the procedure is correct because following it correctly usually gives the correct (independently given) result. Constructivism says: the result is correct because it issues from the correct reasonable and rational procedure correctly followed. (242)

Some details of the account need to be worked out, but let us suppose, for the moment, that Rawls is right in his characterization of Kant's ethical theory: the CI-procedure is how we test our maxims, and it is a constructive procedure. Further, let us accept that Rawls is right about what the fact of reason is. He writes,

> The fact of reason is the fact that, as reasonable beings, we are conscious of the moral law as the supremely authoritative and regulative law for us and in our ordinary moral thought and judgment we recognize it as such. (Rawls 2000, 260)

Let us understand this to mean that it is part of our ordinary moral phenomena that we appeal to some version of the CI-procedure, and that we recognize its authority as supreme. So in this constructivism, what role does the fact of reason play? It plays the small and definitive role of asserting that we do, in fact, rely on the CI-procedure.

I will try to explain. Let us assume that Rawls's constructivist account of Kant's moral theory is the correct account of morality. This would have many implications: some about the metaphysical status of moral claims, some about the epistemology of our moral knowledge, some about our moral obligations, and also some about moral phenomena. The two central claims of Rawls's constructivist interpretation are these. First, its positive aspect: correctly applying the CI-procedure produces valid moral claims with supreme authority. Second, its negative aspect (which follows from the first): nothing else is needed for the validity of those moral claims. Hence, all that is needed to show that morality is real is to appeal to certain facts: namely, that we human beings do apply the CI-procedure, and just as importantly, that we recognize its authority. And the fact of reason asserts just that: we do both those things.

This characterization glosses over many details, but it will do for the moment. I also want to make clear that my understanding of Rawls's account here is somewhat speculative. Rawls is not entirely perspicuous, but I believe that my interpretation of his view is implied by what he says about constructivism and the fact of reason. Here are a few reasons in favour of thinking that Rawls holds such an interpretation, and in going over these reasons, I think we can get a clearer account of his constructivism.

First, Rawls believes that Kant's conception of objectivity is not the realist's one. Rawls writes, "a correct moral judgment is one that conforms to all the relevant criteria of reasonableness and rationality the total force of which is expressed by the way they are combined into the CI-procedure" (244). Rawls also says that satisfying all the relevant criteria, or passing the CI-procedure, is sufficient, that there is no need for any "empirical explanation of how we know it is correct", and that the only explanation we need to know that a moral judgment is correct is that "we have correctly applied the principles of practical reason" (245). What is implied by these claims is that a moral judgment is correct *if and only if* it conforms to the CI-procedure. All that it means for a moral judgment to be true *just is* that

it meets the demands of the Categorical Imperative. But since the question of the reality of morality can be answered by proving the correctness of a given moral judgment, what we need is that a given moral judgment can pass (or even fail to pass) the CI-procedure. This requirement is easy to meet in the abstract. We consider a maxim, subject it to the CI-procedure, and then arrive at a result which tells us whether acting on such a maxim was permissible or not. But morality would remain an abstract fantasy, a concoction of a daydreaming philosopher, if it were too far removed from the mental life of ordinary human beings. It would have no sway on us if we did not, at least tacitly, already recognize and respect the place of morality in our lives. And that is what the fact of reason asserts: that our consciousness of the moral law "is found in our everyday moral thought, feeling, and judgment; and that that law is recognized as authoritative, at least implicitly by ordinary human reason" (Rawls 271).

Second, Rawls says that there are four necessary and sufficient conditions for pure practical reason. They are as follows:

> (1) The content condition: the supreme moral law must have sufficient structure to rule out possible maxims.
>
> (2) The freedom condition: the moral law must be understood as a principle of autonomy.
>
> (3) The fact of reason condition: "our consciousness of the moral law as supremely authoritative for us as reasonable and rational persons must be found in our everyday moral thought, feeling, and judgment; and the moral law must be at least implicitly recognized as such by ordinary human reason."
>
> (4) The motivation condition: "our consciousness of the moral law... can be a sufficient motive for us to act from it" (254-255).

I do not wish to explain all these conditions; rather I want to draw attention to a peculiarity about them. Rawls believes that these four conditions are sufficient for there being pure practical reason. Suppose we take conditions (1), (2), and (4) to be satisfied. These are all claims about what morality entails, but none of them assert that morality is real. Nothing

in (1) and (2) directly requires the reality of morality; rather, they explicate how the moral law must be understood if it applies in our lives. We can likewise think of (4) as stating something conditionally: if the moral law were a valid force in our lives, it can by itself be a sufficient motive for us to act from it. So if we take (1), (2), and (4) as being satisfied, we ought to consider why Rawls thinks satisfying condition (3) is all that is left to do. This may seem perplexing if you think that the true moral law is meant to represent some independent, external reality, because even if (4) were satisfied, it could all just be in our heads, so to speak—we would still need to satisfy a fifth condition, something along the lines of "the moral law does in fact accurately describe some independent, external moral reality." None of the four conditions, either separately or conjointly, guarantee that the Categorical Imperative accurately describes some independent, external moral reality. But if we reconsider Rawls's constructivist interpretation, it becomes easy to see why satisfying (3) is now enough if we already have (1), (2) and (4). There is no need for a fifth condition. If the fundamental moral law is a constructive procedure, then there is no independent moral reality, and no need for one either. Conditions (1) to (4) will thus be sufficient. I hope that these various considerations not only help explain Rawls's view, but also help explain my interpretation of his view.

We can get a different understanding of constructivism and the role the fact of reason plays by considering how together they answer moral scepticism. There are, of course, various kinds of moral scepticisms, but I wish to draw a parallel between some of them to external-world scepticism in epistemology. As I shall understand it, external-world scepticism is the thesis that we cannot know that the external world exists. And the argument for this usually takes the following form. First, we have the set of all our phenomena P. Second, there are a variety of possible circumstances $C1, C2, ...Cn$, each of which can give rise to that same set of phenomena P. (Let us assume further that, apart from the fact that they can all give rise to the same phenomena, there is nothing in common among all these

circumstances.) And, hence, we have the sceptical conclusion: we cannot get from our phenomena P to any specific C with any certainty. Since one of the C's is the hypothesis of the existence of the external world, we cannot prove, with any certainty, the existence of the external world from our phenomena.

There is thus a gap between the set of our phenomena and the actual circumstances that gave rise to our phenomena. One solution to this problem, perhaps a kind of idealism, is to deny that we need to look for anything beyond our phenomena. More specifically, it would assert (1) that the external world is part of our phenomena, and (2) that it makes little sense to think of the external world as independent of our phenomena. Both of these two claims are odd, but we can make some sense of them by thinking of the external world as somehow being constructed from, or being some function of, our phenomena and of nothing else. And thus, this account posits that the setup errs: there isn't our phenomena, on the one hand, and the hypothesis of the existence of the external world, on the other. Rather, since the external world is some function of our phenomena, they're both on the same hand. Even if there remains a variety of possible circumstances, all of which can produce our phenomena, the hypothesis that the external world exists is not one of the possible circumstances. That the external world exists belongs on the side of our phenomena P, and not as one of the circumstances. I raise the suggestion of idealism because it shares some features with moral constructivism.

Moral scepticism as I'm concerned with here is not perfectly analogous with external-world scepticism, because moral scepticism's thesis is not "that we never know what the right thing to do is" (external world scepticism says "we never know what the external world is like"), but "that there are no moral truths". I believe that the analogy is helpful nonetheless, because I think that the main difference is that external world sceptics have no problem thinking that there is an external world while the moral sceptics have a problem thinking that there is an independent moral reality. But both question using our experiences

as a way of getting at something independent. So here is how I think of the moral situation.

First, we have the set of all our moral phenomena *MP*. Second, there are a variety of circumstances that $D1$, $D2$,... Dn, each of which can give rise to the same set of moral phenomena. Let us divide the set of possible circumstances into two camps: (1) those in which there is an objective and independent reality of moral truths, which can give rise to our moral phenomena, and (2) those which can give rise to the phenomena, but there is no objective and independent reality of moral truths. Those who believe that the actual set of circumstances is of type (1) are moral realists; and those who believe that the actual set of circumstances is of type (2) are moral sceptics. Possible contenders for circumstances that fall under (2) might include the situation where our biological instincts are sufficient to bring about such phenomena.

But merely setting it up this way, provided that there are alternative sets of circumstances which can give rise to the same moral phenomena, is enough to engender a weak moral scepticism: we can never know whether our moral obligations are real. There is a gap. We cannot tell which set of circumstances gave rise to our phenomena. But most moral sceptics hold onto its stronger variant—that there are no true moral claims. In effect, they rule out certain sets of circumstances (the ones where objective moral claims obtain) as being causes of our phenomena. This, of course, requires further argumentation.

Many such arguments, I believe, take the following form: It is more metaphysically plausible to hold that there is no objective and independent reality of moral truths than otherwise.[3] And since presupposing that there is an objective and independent reality of moral truths is unnecessary for the production of our moral phenomena, we ought to infer that there is no objective and independent reality of moral truths. The variety of arguments for moral scepticism arise from the variety of reasons for thinking that it is more metaphysically plausible to deny the objective and independent reality of moral truths. For instance,

[3]Mackie (1977) and Harman (1977) both make similar arguments to this effect.

some may hold that denying its existence fits better with a scientific view of the world, while others may insist that it is more ontologically parsimonious.

And on the positive side, the strongest reason for believing in the existence of an independent moral reality, which I think a fairly compelling one, is that denying the existence of an independent moral reality gives up too much. It would mean that many of our strongest moral convictions—that the Holocaust was an abomination, that it is wrong to murder innocents for fun or profit—are not, strictly speaking, true; instead, it would mean that they are false or that our convictions are not beliefs at all. It would imply that these convictions are, in the end, only phenomena to be explained.

Those compelled towards moral realism must somehow argue that the sets of circumstances that do not require the existence of an independent moral reality cannot produce all the relevant moral phenomena. This may be difficult, because it is hard to rule out other possible explanations a priori. The only option seems to be to establish the existence of an independent moral reality by showing that it is a necessary condition for the possibility of our moral phenomena.

But there is another option. We can argue that the whole of our moral life is much more than, and cannot be reduced to, a set of mere phenomena. That is, the setup is faulty. One way of doing this is to defend the idea that there is some genuine moral feature that does not need to be grounded in an independent moral reality. Once we think that the validity of a moral feature requires that it reflect or represent some independent moral reality, the possibility of moral scepticism arises because there is always some alternative explanation for that feature. So the idea here is to claim that the validity of some moral feature does not depend on any external reality—for instance, that our being obligated to act in particular ways (notice the lack of an abstract noun) is sufficient enough on its own for our actually having an obligation (notice the presence of an abstract noun), and that there need not be any objective and independent entity we can point to that we can call an "obligation".

Likewise, we do not need there to be any objective and independent entity that we can call "the right" or "the good" and so on.

Rawls's constructivist interpretation of Kant's moral theory, I believe, offers one way of realizing this option. Recall that it is constructivism's ambition to assert that the elements of our moral life need not be grounded in any independent moral reality. On the constructivist account, many, if not all, of our moral judgments will be constructed from a procedure. That a particular moral judgment is constructed correctly by the CI-procedure is both a necessary and sufficient condition for a particular moral judgment being true. And so there is no need for any particular moral judgment to somehow reflect an independent and external reality, nor is there any reason to look for one. The correct construction is all that is required for a particular moral judgment to be true.

And so if constructivism is true, all that is needed to justify our moral life is to show that we have one. Namely, that we do, in fact, have moral judgments. But it would be odd to look for an argument for this claim, as if one could find plausible premises to this desired conclusion. It is like trying to provide an argument that we have experiences. Rather, we should simply assert that we do in fact have experiences, and that we do in fact make moral judgments. That is what, I think, the fact of reason does, on Rawls's account: it asserts that we have a moral life, and that moral considerations factor into our practical deliberation.

What makes Rawls's constructivism Kantian is its content: the constructive procedure embodies Kant's Categorical Imperative. On Rawls's account, Kant's argument for the claim that the CI-procedure is the relevant constructive procedure is given by an analysis of the concepts of the motive of duty and a good will. This was already done in the first two chapters of the Groundwork, which I do not here attempt to defend.

Before I offer my own criticisms of Rawls's account, there's an objection that I would like to consider.

Objection: Constructivism seems to allow for moral relativism. If correct moral judg-

ments are the results of a process of construction, then it seems that the only correct moral judgments we have are actual instances of construction. We could only distinguish the correct moral judgments from the incorrect ones because the correct ones adhere more closely to the CI-procedure. But being the closest instances is no guarantee that they are close at all. For instance, the closest adherence to the CI-procedure, at some point in the past, could condone the institution of slavery. And then constructivism would force us to say that slavery was once morally acceptable. I can see how one might avoid this relativism, if constructivism asserted that only a perfectly correct following of the procedure produces valid moral claims. But if this is what constructivism demands, this would have an odd result. Since it's possible that with respect to a particular course of action, no one has yet to follow the CI-procedure perfectly, then there is not yet any valid moral claim regarding that particular course of action because no act of construction has occurred. But if you want to insist that there is already a valid moral claim, as if the CI-procedure by itself determined what was right or wrong, before anyone has actually correctly followed the CI-procedure, then isn't that already a kind of moral realism?

Reply: Constructivism does allow for different ways of realizing the set of valid moral claims. But this does not imply relativism. Some putative moral claims will be ruled out from the beginning. As in contractarian political theories, various ways of realizing the state will be possible, but some political institutions will be ruled out from the beginning. How a society determines the set of valid political claims depends on empirical features of that society; in particular, it depends on what its various people desire and value. Certain rules or laws will only have legitimacy because they are the outcome of following a procedure. Incorrectly following the procedure will not result in a legitimate law, because it is no different from not following the procedure at all.

And even though certain claims will be ruled out from the beginning, this does not necessitate realism. Rawls makes an analogy with constructivism in mathematics. He

writes,

> the idea is to formulate a procedural representation in which, as far as possible, all the relevant criteria of correct reasoning—moral or mathematical—are exhibited and open to view. The idea is that judgments are valid and sound if they result from going through the correct procedure and rely only on true premises.... In arithmetic, the procedure expresses how the natural numbers are generated from the basic concept of a unit, each number from the preceding.... The procedure exhibits the basic properties that ground facts about numbers, so propositions about numbers that are correctly derived from it are correct. (238-9)

So in constructivism in mathematics, no one need correctly have proved a given theorem in order for it to be valid or sound. All that matters is that the theorem *can* be derived from the axioms. Likewise in moral constructivism, if correctly following the CI-procedure would show that a particular maxim is impermissible, then it is impermissible even if nobody has correctly followed it. But as in constructivist mathematics, this does not mean that there is an independent mathematical reality which mathematicians, in their practice of mathematical deliberation, are attempting to match. And so too in moral constructivism. Perfect reasoning would produce valid moral claims, and when we are engaging in moral deliberation, we are just trying to reason correctly. Even though some moral claims must be valid or invalid because the procedure of construction doesn't allow everything, this doesn't imply realism, because it doesn't presume an independent moral reality.

So I think this charge of relativism can be answered, but there's a deeper charge of relativism, whose answer reveals a certain difficulty with Rawls's account of Kant's constructivism. The deeper charge is this: if moral claims are constructed, it seems as if we can construct anything. And so what is right or wrong will depend entirely on what we choose to construct and we can choose anything. Thus, we have moral relativism.

The answer to this objection, of course, is that not anything *can* be constructed, because not everything *is* constructed. Most importantly, the process of construction is not up

to us. But this answer is revealing. Rawls often seems to be trying to articulate a *metaethical* position, one that is distinct from moral realism or moral sentimentalism, but at the same time, his constructivism already assumes substantive normative claims. In particular, it assumes substantive normative claims about the value of equality, freedom, reasonableness, and persons. The CI-procedure embodies such values, and for Rawls, it is not itself constructed.

Rawls says that the CI-procedure "is simply laid out" (239). This is obscure, but I believe his point here concerns its justification, or more properly, it concerns what constitutes our legitimate employment of the CI-procedure. According to Rawls, the CI-procedure is based on "the conception of free and equal persons as reasonable and rational" (240). This, however, does not mean that we first have a conception of persons, as free and equal, as reasonable and rational, and then derive our CI-procedure from these conceptions. (Rawls uses the term "rationality" here to refer to a capacity to organize one's interests and make decisions which advance them; whereas "reasonableness" here refers to a willingness "to listen to and consider the reasons offered by others" (164).) Rather, the CI-procedure somehow embodies or reflects these conceptions, and we can discern the moral conception of a person from our practices of moral deliberation. Rawls says that these conceptions (of person and society) "are elicited from our moral experience and reflection and from what is involved in our being able to work through the CI-procedure and to act from the moral law as it applies to us" (240). We can see how this works. The CI-procedure asks us, in effect, to make sure our reasons for acting could also be reasons for anybody to act in similar ways. It demands of us to think of ourselves as members of a community of people, each free to pursue her own interests and plan her own life, and all equal with no one having any intrinsic authority over others, or special rules that only apply to a subset. And this goes some way to answering the question of its justification. There is a kind of coherence between our CI-procedure and our conceptions of persons and society. In other words, the

CI-procedure and our conceptions of persons, equality, and freedom are the results of a reflective equilibrium.

Even though this coherentist justification for these particular substantive normative judgments—the ones about persons, equality, freedom, etc.—may suffice for answering normative questions about our values, it doesn't tell us much about their metaphysical or semantic nature. Since, as Rawls says, they're not constructed, there must be some other account of their nature. But Rawls doesn't provide any another account for these claims. And if he is silent on the matter, then, as Sharon Street points out, "they are in principle compatible with any number of competing metaethical views" (217). (She's actually writing about Rawls's political constructivism rather than his account of Kant's moral theory, but the point remains.) We could be realists, or expressivists, about our basic normative judgments concerning equality and freedom, and at the same time, be constructivists about the rest of our moral judgments, which we take to issue from these basic normative judgments. But this implies that Rawls's account of Kant's constructivism is not a genuine alternative to realism or to expressivism. And insofar as it fails to do that, it fails to live up to Rawls's project of articulating a unique metaethical view.

But not all is lost. This silence about the metaethical status of the basic normative judgments doesn't imply that Rawls's constructivism is untenable or that it doesn't have metaethical implications. Indeed, it seems perfectly coherent to hold that many of our moral judgments are constructed, and hold that our basic moral judgments, ones concerning equality and freedom of persons, may not be. And it has metaethical implications too. After all, if this kind of constructivism were true, many of our moral judgments would be constructed. We would understand our epistemic situation with respect to these judgments, and they would have a palatable metaphysical status. But with respect to both their metaphysical nature and to our epistemic access to them, our understanding would be limited, because they would, at bottom, depend on how we understand our more basic normative

judgments.

That constructivism fails to be a distinct metaethical position is not necessarily a problem for it. But as an interpretation of Kantian ethics, it runs contrary to the standard interpretation. In particular, the standard view holds that Kant tried to derive our moral obligations from the concept of rational agency (or more accurately, he tried to show that in virtue of our being rational agents, we have certain moral obligations or that we are committed to certain rational requirements that entail substantive moral results). So the standard view of Kant does not assume that we value certain conceptions of persons as free and equal prior to our having the Categorical Imperative and the particular moral judgments it issues. Under the standard view, it is in virtue of our being rational agents alone, and not in combination with other values, that we are committed to the Categorical Imperative. Under Rawls's view, this does not seem to be the case. The two views—the standard and Rawls's—could be reconciled if we can somehow articulate how we, in virtue of inhabiting the perspective of rational and practical agents, must value equality and freedom. But Rawls offers no such account.

There's a separate and more important difficulty with Rawls's account, in particular with his reading of the fact of reason. Recall that Rawls says that the fact of reason is "the fact that, as reasonable beings, we are conscious of the moral law as the supremely authoritative and regulative law and in our ordinary moral thought and judgment we recognize it as such" (Rawls 260). This is ambiguous between two very different claims.

> (1) We are conscious of moral considerations as being supremely authoritative and regulative, and in ordinary moral thought and judgment we recognize morality as being so.
>
> (2) We are conscious of the Categorical Imperative (or the CI-procedure) as being supremely authoritative and regulative, and in ordinary moral thought and judgment we recognize the Categorical Imperative (or the CI-procedure) as being so.

The first says that if we recognize an action as moral, then we recognize it as being authoritative. It's a view consonant with many commonly held views about morality; for instance, it respects a common version of motivational internalism, which says that if an agent judges an act to be right, then either she is motivated to perform that act or she is practically irrational. It's also consonant with many standard views about Kant's ethics, especially on the value of acting from the motive of duty. It's also widely held to be true that we do ordinarily take moral reasons to trump all other practical reasons.

And so if the fact of reason is understood this way, then it seems to be true and it is relatively unobjectionable. It is thus tempting to read the fact of reason this way. But read this way, the fact of reason is consistent with a variety of different normative ethical theories. One could, for instance, be a utilitarian and consistently hold that it is a fact that we recognize that moral considerations outweigh other practical considerations. In fact, even a moral sceptic could consistently hold that it is true that we human beings ordinarily think moral reasons are authoritative so long as she denied that they actually were authoritative. And I think it is even possible for a moral egoist to hold that moral reasons trump other practical reasons. There seems to be nothing distinctively Kantian about the fact of reason understood as (1). If that's right, then it's difficult to see how it could show to us our transcendental freedom, as Kant seems to think it does.

If, on the other hand, we read the fact of reason as (2), we can achieve some of Kant's objectives. Since the Categorical Imperative gives us an imperative that does not depend on any ends we happen to have or desires we want to satisfy, once we realize that we can act in accordance with it, we must regard ourselves as not being tied to our empirical selves. We can thus regard ourselves as being free. And thus, we can see how reading the fact of reason as (2) can help us understand Kant's later objective of showing that the moral law reveals to us our freedom. Moreover, the fact of reason is ineluctably Kantian. It would be of no use to any other normative ethical theory to appeal to such a fact.

But unfortunately, read as (2), the fact of reason is less obviously true. We are *not* all conscious of the Categorical Imperative as being supremely authoritative, nor do we recognize it as being so in ordinary moral thought and judgment. Read this way, it would be hard to convince a non-Kantian that the fact of reason was true. Utilitarians, Aristotelians, moral egoists, and moral sceptics would easily deny it. The fact of reason, read as (2), is highly contentious, and it hardly seems to be a fact at all. And if it is this contentious, we have grounds to be sceptical of its capacity to serve as the essential element in a justification of morality.

We could try collapsing (1) and (2), and perhaps explore the possibility that by the time Kant gets to discussing the justification of morality in the second *Critique*, he has already argued, and thus already assumes, that morality must look a certain way: namely, that the Categorical Imperative is the moral principle we all recognize, even if we do not know it. But to the extent that we succeed in doing this, we are still left with the most serious problems with reading it as (2): the fact of reason becomes less plausible.

In any case, I do not believe that this problem is particular to Rawls. Rather I believe it is particular to Kant's theory, but Rawls's account fails to address this worry. I shall explore it further in the next chapter when I articulate what I take the fact of reason to be.

3.4 Allison's interpretation

Allison's view has a lot in common with Rawls's, but there are also significant differences. The most important similarity is that Allison would reject, along with Rawls, the setup which leads so easily to moral scepticism. Rawls's constructivism offers a way of making sense of rejecting that setup, by reframing how we think about the elements of our moral life. The important difference is that Allison doesn't endorse constructivism. But Allison doesn't explicitly offer an alternative way of thinking of our epistemic situation regarding

morality. I will, however, argue that his account of Kant's view of the structure of *moral* experience is best understood as analogous to Kant's view of the structure of *ordinary* experience. (Kant does not make the analogy, but Allison does. I explore this analogy further, because I think it proves useful in helping us understand Allison's interpretation.) So the extent to which transcendental idealism reshapes our epistemic situation so as to answer sceptical problems, Kant's understanding of our moral life also reshapes the epistemic situation concerning moral scepticism. In other words, there is something like a transcendental idealist solution to moral scepticism. There are, however, significant differences between our theoretical, scientific knowledge of the world and our moral, practical life, which help explain the 'unargued' justification for the fact of reason, but which, at the same time, make this account unsatisfying.

I argued above that moral scepticism, and especially a weak moral scepticism, becomes easy to accept once we think that the validity of our moral experiences depends on accurately reflecting an external reality. Constructivism offers a way of making sense of the validity of our moral experiences while not requiring that it depends on reflecting some external reality. Constructivism rejects the way moral sceptics describe the character of our moral experiences and the standard of validity that moral sceptics demand.

I believe that Allison also rejects the assumptions that lead to moral scepticism, but he is less clear than Rawls is about the structure of his alternative. Much of the reason for his lack of clarity on this issue is that he is largely uninterested in defending morality against moral sceptics. But Allison does, I believe, attribute to Kant an alternative account, which takes a parallel structure to Kant's rejection of epistemological scepticism.

The important clue to my interpretation lies in Allison's assessment of why Kant thinks there is no deduction of the moral law. According to Allison, if there were a deduction for the moral law, it would parallel the transcendental deduction of the categories in the first *Critique*. He says that it would presumably take the following form: we would take note

of our moral experiences, perhaps identify some characteristic, and then argue that the moral law is a necessary condition for our having such experiences (235). Instead, Allison thinks that Kant holds that the moral law is given to us as part of our moral experience, it is "an ingredient given in it" (235). That Allison conceives of this as the most significant difference between a 'proof' of the moral law and a deduction of the categories is telling. This suggests that we can learn much by comparing the two (between the structure of our knowledge of the external world and our moral and practical knowledge) and see what accounts for the difference. Since Kant's transcendental idealism serves as an answer to epistemic scepticism, we should explore the parallel in the moral case to see what kind of answers we might have.

This parallel would be straightforward if there was a standard or obvious interpretation of the transcendental deduction. There isn't. But we don't need to understand the transcendental deduction in its full details. Looking at its structure, which is fairly uncontroversial, should be sufficient for my purposes. So its structure is this: Kant first tries to characterize the nature of our experience in such a way that it is relatively uncontentious. And then, he tries to examine that experience and to prove that certain claims would have to be true in order for us to have such experience. In particular, he wants to establish that we can legitimately apply certain a priori concepts to our experience.

Different interpretations of the transcendental deduction will give different accounts of how Kant characterizes our experience, how uncontroversial his characterization of our experience is, what it means to apply a priori concepts to our intuitions, etc. For the most part, these issues are largely irrelevant here. What I'm interested in is the conditional nature of the conclusion. We can legitimately use a priori concepts only because we have certain kinds of experiences. For instance, we can legitimately appeal to a principle of causality because for Kant our phenomenal experience would be different without it, perhaps so different that it would be utterly unrecognizable as phenomenal experience.

But it is important to note that the transcendental deduction, for Kant, is not the proof of an external reality. The categories (or the a priori concepts), after all, are not external entities. They are subjective in a sense—in the sense that we cannot have thoughts about our intuitions, or even experience them in the way we do, without placing those intuitions under certain concepts.

But there is no proof that these concepts and categories help organize things as noumena, as things outside of the way we experience them. There is no way we can get there. This is the "idealist" part of transcendental idealism; it's entirely about our experiences. The transcendental deduction, then, is an attempt to *justify* our use of such concepts and categories, rather than an attempt to prove that they have external and independent reality.

Notice that the criterion of success has changed. Kant does not try to prove, for instance, that the law of causality—that every event has a cause—is true for things in themselves, but rather, tries to prove that we can legitimately employ such a concept. And the legitimacy of the categories depends, if the argument works, on the premise that the character of our experience would be utterly different if they didn't play a role in organizing it. But this amounts to a weak legitimacy; it is conditional upon the structure of our experience. And though Kant is only trying to show that such concepts have no less legitimacy than our experiences, they would also fail to have any more legitimacy.

But let us turn to see how this helps explain Kant's justification for the moral law. Keeping the parallel with the transcendental deduction in mind, we can guess that Kant would likely think that it is impossible, in principle, to establish the independent reality of the moral law, in the same way it would be impossible to establish that the law of causality holds among noumena. Sebastian Gardner writes, "A philosophical deduction is required whenever we seek to answer a question of rightfulness or justification, as opposed to a question of fact" (136). So with respect to morality, I believe that Kant, in Allison's interpretation, is trying to justify our "employment" of the moral law (B117/A85), rather

than its independent reality.

Kant's crucial claim in the second *Critique* is that there is no deduction of the moral law—that is, there is no argument from moral experience to the moral law as its necessary precondition. But then what takes the place of a deduction? According to Allison, Kant's argument in the second *Critique* is that the moral law is simply given to us in our moral experience. Allison writes, "the moral law is not so much a presupposition of experience as an ingredient given in it" (235).

What does this imply? In the case of the transcendental deduction, the legitimacy of our employing the categories was entirely parasitic on the weight we give to the nature of our experience. That is, the categories stand and fall with experience because the two are intimately connected. But if the moral law is an ingredient of our experience, then its legitimacy is just that of moral experience. It stands and falls with moral experience, not because it is intimately connected with it, but because it is a part of it.

If, say, the concept of causality were given in experience rather than a presupposition of it, then doubting its legitimate employment is the same thing as doubting the nature of experience. And I believe that this is how Allison reads Kant regarding the moral law: to doubt its legitimate employment is to doubt the character of our moral experience.

As in the case of the transcendental deduction, this only makes sense if we accept a criterion of legitimacy weaker than the usual one. We must abandon asking the moral law to be reflective of things outside of experience in the same way we should abandon asking the law of causality to be true of things outside experience. The assumptions for the possibility of moral scepticism are mistaken because we need not, nor can we, look for the external reality that this experience supposedly reflects.

I believe that all this lies behind Allison's interpretation, and it explains why he is unconcerned with articulating an alternative metaphysical view like constructivism. Kant already has one: transcendental idealism. This explains the second disjunct of Allison's

remark that Kant's strategy "can hardly be expected to convince someone who rejects either the account of morality as based on a categorical imperative or transcendental idealism" (230).

What we need to understand now is exactly how Kant thinks of the character of moral experience and to understand how the nature of our moral experience can justify a normative claim. As I understand it, the crucial characteristic of our moral experience is the fact of reason. He takes "the fact of reason" to refer to the fact that, in our practical deliberation, we make moral considerations—we test our reasons for acting with the Categorical Imperative (the FUL formulation). Allison writes:

> [T]he consciousness attributed to 'every natural human reason' is of particular moral constraints as they arise in the process of practical deliberation, with the law serving as the guiding rule (decision procedure) actually governing such deliberation. Moreover, the latter is the law in its 'typified' form, the rule of judgment: 'Ask yourself whether, if the action which you propose should take place by a law of nature of which you yourself were a part, you could regard it as possible through your will.' (233)

Allison doesn't believe that we all have an explicit and conscious awareness of the moral law as a formal principle that we actively and consciously take as our guide. Rather, the idea is that whenever we judge an act to be good or bad, right or wrong, we subject the act to this rule, or something like it, but this does not necessarily mean that we are aware of our doing so.

So two important questions: The first important question is this: do we, in fact, appeal to the Categorical Imperative in its "typified" form? This is a difficult question to answer, because, according to Allison, Kant thinks it is simply true. After all, Kant says, in the section "Of the typic of pure practical reason" that "everyone does, in fact, appraise actions as morally good or evil by this rule" (5: 69). Allison says it is a "brute given, which cannot be derived from any higher principles or from a reflection on the nature of rational agency" (Allison 233). I'm not sure how to prove a brute given, if it can be done at all. But we

cannot simply assert that a claim is a brute given, and then declare that no more can be said about it. If we could, we would be warranted in making whatever assertion we wish, so long as we also asserted that it was a brute given. We have to say something to the objectors, and the objectors can include Kantian ethicists because even though they think the Categorical Imperative is true, they don't need to think that it is obvious at all.

The best I can do to argue for the presence of the Categorical Imperative in our ordinary practical deliberation is to explicate its meaning in plain terms, and hope that in its articulation an objector will realize that she too believes it. For example, if someone claimed that he didn't believe in modus ponens, what can one do but explain as clearly as one can what it means? How else do we argue for a brute given?

And for this task, the most promising and most plausible explication of the Categorical Imperative is that it ought to be understood as demanding of us to make our reasons universal; that is, we take a reason for an action to be legitimate only if we allow that it would be legitimate for anybody to perform a similar action in a similar position. It would hardly count as a reason otherwise. This is a familiar argument in Kantian ethical thought. As Allison puts it, "the universalizability of one's intention, maxim, or plan of action seems to be presupposed as a condition of the possibility of justifying one's action, even when this justification does not take an explicitly moral form" (205). In order for me to count a reason for an action as good, I must think that it would be a good reason for anybody in a similar situation. It doesn't seem to count as a reason otherwise. This argument, of course, is not without its problems. In particular, this description of the requirement of the universalizability of reasons doesn't exactly match Kant's formulations of the Categorical Imperative.

If our aim is to make the Categorical Imperative, and not some weakened version of it, palatable, I think we would do best by articulating its typified version, the rule, Kant believes, by which we test whether an action is morally acceptable or not. And it says that

when we consider an action, we imagine living in a world where the policy that underlies that action is, in fact, a universal law of nature, and so everybody always acts in accordance with that policy, and if we could and would will such a world, then such an action is permissible. Kant doesn't say that we always do this, but whenever we consider whether an action is moral or not, we do. But do we really imagine worlds where our maxims are universal laws of nature in order to judge whether our actions right or wrong? Perhaps we do, but this claim—that this is what we do—is far from obviously true.

It seems that by explicating the Categorical Imperative, we have not made it more easily acceptable. In this case, it is unlike *modus ponens*. The weakened form of the Categorical Imperative might be easy to accept, but the typified version is not obvious at all. And unfortunately, Kant's view seems to need the strong version.

The second important question is this: how does placing the moral law within our experience make employing the Categorical Imperative legitimate? I argued above that this requires a criterion of legitimacy different from one that we use for our knowledge claims. In particular, we have to claim that moral obligations can be legitimate without reflecting any external reality. There are a lot of good reasons to think this. In particular, it is just odd to think that any purely descriptive account of the universe, or some part of the universe, by itself, could tell us how things ought to be. Yet that is what the moral sceptics demand: they set up a criterion for justification or legitimacy which would be senseless to satisfy. That is, moral sceptics ask of anyone who defends morality to do two things: to show that there is an objective, mind-independent moral reality, and to show that our moral judgments can be the result of having epistemic access to that reality. But Kantians rarely believe that there is an objective, mind-independent moral reality, so they reject the sceptic's request as pointless. Also, some non-Kantian moral realists think that no purely descriptive account of a part of the universe will ever entail claims about how we ought to behave. For instance, Russ Shafer-Landau, a prominent moral realist, suspects that moral principles have no

independent causal power. But according to him, that says nothing about whether moral principles exist or not, because moral principles are normative and prescriptive, and hence outside the realm of scientific investigation (228). So, what new criterion of legitimacy could we have for the moral law, especially if it's just given to us as part of our experience? As far as I understand Allison's view, if we can explain why this fact is a fact about *reason*, and not a fact about our inclinations or human psychology, then we have some justification.

According to Allison, the explanation is already given in the *Groundwork*. Once we realize that the principle we abide by as a matter of fact does not appeal to or depend on our desires or inclinations or any other empirical condition, we have to conclude that it is a product of pure reason. Allison says that "it addresses us in our capacity as rational agents and with a claim to universality and necessity that makes no concessions to our sensuous nature and no reference to empirical conditions" (236). The typified form of the Categorical Imperative demands of us to imagine a world where the maxims that underlie our actions are universal laws of nature and only if we could and would will such a world would our actions be permissible. If we followed such a principle, there is no desire or empirical inclination being satisfied. And since it doesn't depend on any empirical feature, then it must be something only pure reason could deliver. And since it's a product of pure reason, it is legitimate.

I think the answers to these two questions help motivate Allison's claim that the justification of the moral law "has already been attained, implicitly at least, in and through the analysis of the nature of morality and its principle" (236). What we had to understand was why an analysis of morality was sufficient. If the issue of its legitimacy is resolved insofar as we seek nothing beyond our moral experience, and if the analysis of our moral experience yields the Categorical Imperative (as the first two parts of *Groundwork* attempt to show), then we have the moral law insofar as our experience takes on a particular character, which is what the fact of reason supplies.

In summary, Allison believes the following: it is not the case that we have to justify the Categorical Imperative, via a deduction in which we would argue that it is a precondition of our moral experiences. Rather, it is simply given to us. We human beings invariably appeal to the typified form of the Categorical Imperative in our practical deliberation. And this is the fact of reason. The only question that is possibly at issue is whether we are justified in using it, and that criterion of justification cannot be whether it reflects an independent and external reality. Instead, all we need to do is show that it is a product of pure reason. And we can show that it is a fact of *reason*, because the Categorical Imperative appeals to us in our reason and not to any contingent fact about us, either in general or in particular.

Moreover, there is no transcendental deduction for the moral law. When it came to the categories, there was a transcendental deduction, because there are two faculties—sensibility and understanding—both of which are needed to explain the character of our experience. Practical reason isn't the combination of two faculties. Unlike in the case of an experience of a physical object, we cannot separate, or even distinguish, the concept of a particular moral judgment from its content or vice versa in the way we can conceptually distinguish the barest sense experience and the categories by which we understood it. And hence, no transcendental deduction is available. The moral law is just given to us. This, in sum, seems to be Allison's view.

But there are problems with this view. I've already mentioned one: the fact of reason, as Allison conceives of it, is far from obvious.

A related problem is this. If our moral experience contains the moral law, as Kant would have it, it is much richer than the moral experience our moral sceptic conceives of. I imagine the sceptic would simply object to the character of our moral experience. The transcendental deduction of the categories in the first *Critique*, regardless of which interpretation, seems to have had a better starting place than here in the moral case, because there, the character of experience was meant to be rather slender such that even sceptics, or

at least most sceptics, could accept it. Kant tried to point to neglected but uncontroversial features of our sensory experience and then argue to their necessary conditions. Here in the moral case, things seem to be more questionable from the get-go. Our moral consciousness is hardly taken to be slender. If the fact of reason is to be believed, it already contains the moral law in the form of the Categorical Imperative. And this is hardly obvious. Sceptics can easily deny that we, in fact, do anything like appeal to the typified form of the Categorical Imperative, and instead assert that it is highly unlikely and implausible that when we judge an action's moral permissibility, we imagine a world where the maxim that underlies it has become a universal law of nature, and we ask whether we could or would will such a world. And despite what Kantians may think, I suspect most philosophers would side with the sceptics on this issue.

There is, however, another nagging worry. How is all this possible? How is it possible that pure reason can give us a principle that we take as ultimately guiding? How can reason give us this constraint on our practical deliberation? In other words, how can pure reason be practical?

According to Allison, we can have no answer to such questions. This is the point in which Kant said "human insight is at an end as soon as we have arrived at basic powers or faculties" (5: 46–47). This is because we are now asking about things in themselves. All we can ever know is limited to our experience, and to their necessary conditions. But we cannot know so much about things in themselves such that we can explain *how* they ground our experiences. Like in the epistemic case, we know some things about noumena—in particular, we know that they must exist, but we cannot know what properties they have or what laws they obey. We know next to nothing about them. In the moral case, the moral law is a part of our experience, and there is little to use to argue for the necessary conditions of our moral experience. Even if we could argue to some necessary conditions, we have no way of getting all the necessary conditions; we would still be left without any adequate

explanation of how pure reason can be practical. We may be able to conjecture about how things in themselves ground our moral experience, but it can only be conjectural. We would be doing the kind of metaphysics that Kant thought was impossible.

But this answer is unsatisfying. After all, we have no grounds to accept this quietism about the practicality of pure reason, unless we already deny that we can know anything about it because we already accept some form of transcendental idealism. The inexplicability of the practicality of pure reason seems to amount to nothing more than stipulation. We are told that we cannot ask how pure reason is practical because explanations come to an end when we ask about the powers of our primary faculties. We are told that experience is as rich as it is and that it ineluctably takes on particular characteristics. If we don't accept transcendental idealism, then we might take it as a condition of an adequate theory that it explains how it is that we can be moral. Allison's theory seems to end the inquiry too early.

3.5 Conclusion

Beck, Rawls, and Allison have all offered very interesting and insightful accounts of Kant's account of the fact of reason. They all illuminate different aspects of Kant's thought, and they all offer innovative ways of thinking about how to ground Kantian ethics. But as rich as their accounts are, they seem to suffer from certain difficulties. In the next chapter, I will try to offer an account that meets some of these difficulties.

Chapter 4

The Fact of Reason and the Justification of Morality

4.1 Introduction

In the *Groundwork*, Kant attempted to justify morality on the grounds that it follows from our rationality. This consisted in establishing two basic claims. First, if we are theoretically rational, then we must also be practically rational.[1] And second, if we are practically rational, then we stand under the fundamental principle of morality. Kant abandoned this project when he came to write the second *Critique*. Because Kant was never explicit about what he rejected from the *Groundwork*, it is not clear what step he thought in it mistaken. We do best by trying to give plausible accounts of both arguments, and see what the new argument fulfills that the older one does not. I believe that Kant came to reject the idea that our theoretical reason shows our practical reason. And I also believe that Kant also came to reject the idea that our practical rationality, by itself, could entail our obligation to the moral law; I believe it is this last step that the fact of reason argument replaces.

But if the fact of reason suffices to prove the reality of morality, can it show that we are

[1]The distinction between theoretical reason and practical reason goes as far back as antiquity. Roughly speaking, to be theoretically rational is to have reasons influence one's beliefs. And to be practically rational is to have reasons influence one's actions, plans, or intentions. Kant actually thinks that theoretical reason and practical reason are "one and the same reason", distinguished only in its application (G 391).

practically rational? Can it show that we are *theoretically* rational? We can answer these two questions in three ways: (a) yes to both (b) yes to the first, but no to the second, or (c) no to both. Each option is less ambitious than the one before it. I will be assuming our theoretical and practical rationality when I argue for the fact of reason, and so I will answer 'no to both'. But I can see being tempted to answer (a), because if we do have the reality of morality, both practical and theoretical rationality logically follow. Consider the following analogy: Suppose I try to prove to you that my hand exists. Depending on what your doubts are and what I'm trying to do, I may already be assuming the existence of the external world. For instance, you may think that I don't have a hand because you heard I lost it in the war. Proving to you that my hand exists might consist in merely meeting you and physically presenting to you both of my hands. But the existence of my hand entails the existence of the external world, because no external world means no hands. So a question: is showing my hands to you a proof of the external world? Not necessarily. What I have done instead is assume the existence of the external world, and then tried to meet the other conditions required as evidence for the existence of my hand. And this is what I intend to do here regarding the fact of reason. Even though the fact of reason entails the existence of morality, and that entails our theoretical and practical rationality, I do not take the fact of reason as being able to show that it does in a non question-begging way.[2]

Above, I attributed a dilemma to John Rawls's understanding of the fact of reason. The dilemma arose from an ambiguity in his formulation of the fact of reason. He wrote that the fact of reason refers to "the fact that, as reasonable beings, we are conscious of the moral law as the supremely authoritative and regulative law and in our ordinary moral thought and judgment we recognize it as such" (260). But it is not clear what is meant by "the moral law". Either this means moral considerations in general or it means the

[2]One might see a parallel between my argument here and Moore's famous argument in "Proof of an External World". But I don't intend my remarks above to either impugn or support Moore's argument in any way.

Categorical Imperative. Rawls's claim is thus ambiguous between being conscious of moral considerations as being authoritative and being conscious of the Categorical Imperative (perhaps as the formula of the universal law of nature) as being authoritative. Taking the first interpretation, it is fairly plausible to claim that moral considerations have a kind of authority in our recognized moral consciousness. The fact of reason, understood this way, is plausible. But understood this way, the fact of reason has relatively little that is particularly Kantian. The fact of reason is, on this interpretation, neutral among various normative ethical theories. And it is unclear how it shows to us our transcendental freedom. On the second interpretation, the fact of reason is clearly Kantian; it would assert that we are conscious of the Categorical Imperative as authoritative. Other normative ethical theories will be ruled out. But on this interpretation we lose the virtue of its plausibility. Many philosophers outrightly deny the Categorical Imperative, and it is simply unconvincing to argue that the Categorical Imperative is part of our recognized shared moral consciousness. On this account, the fact of reason looks implausible.

Henry Allison's view avoids the dilemma by doing two things: he makes the fact of reason about what we are implicitly aware of rather than something we are explicitly aware of, and he makes it clear that it is the Categorical Imperative, rather than mere moral concerns, which is implicit in our behaviour. These are both moves I also wish to make. But this suffers from a difficulty similar to the one that faces Rawls's interpretation: it is not obviously true. While it may be plausible that a weakened version of the Categorical Imperative is exhibited in our moral practices and behaviour, it is less plausible that the formula of universal law or the formula of the universal law of nature is. I believe that this problem is unavoidable, and I will argue that to maintain a Kantian ethical view, we must endorse a weakened version of the Categorical Imperative until we have good reason to find a fuller, stronger version being expressed in our behaviour.

The differences between Allison's account and mine turn out to be rather few. His in-

terests (that of trying to give a plausible interpretation of Kant's theory of freedom) differ from mine (that of trying to give a plausible Kantian justification of morality), and that accounts for much of the difference in approach and subject choices. There is, however, one important difference between Allison's account and mine. Even though Allison understands the fact of reason to be an implicit recognition of the Categorical Imperative as a guiding practical rule, he doesn't draw the implications this has for Kant's moral theory. Allison places the fact of reason only within the context of what we do when we try to think about what morality requires of us (233). But how Allison understands the interpretation of the fact of reason has broad implications. The fact of reason understood as an implicit awareness means that the Categorical Imperative can figure into our practical deliberations without our necessarily being aware of its influence on us. It would thus be possible for us to be motivated by the Categorical Imperative without even believing that we are acting for moral reasons. This is not a trivial implication, because it bears on the question of how plausible the fact of reason is and, more importantly, on the question of what has moral value.

It is relatively unnoticed but there is a deep connection between Kant's idea of the fact of reason and his concern for the value of acting from the motive of duty. In fact, I will argue that the fact of reason is best understood as the fact that we *can* act from the motive of duty. Though this claim seems never to have been made, I think it is a natural conclusion to draw, and I will argue for it. I also need to say something about the nature of morality, such that the fact of reason can serve as the justification for morality.

Kant believes that the fact of reason shows us our freedom.[3] The most natural reading of that argument for this claim goes something like this. Occasionally, we are forced to

[3]Kant writes "the moral law... serves as the principle of the deduction of an inscrutable faculty which no experience could prove but which speculative reason had to assume as at least possible..., namely the faculty of freedom, of which the moral law, which itself has no need of justifying grounds, proves not only the possibility but the reality in beings who cognize this law as binding upon them" (CPrR 5: 47).

make a decision between an option that serves our self-interest and one that serves morality. The fact of reason appears to refer to the fact that we can make the moral choice instead of the self-interested one. That we can make such a choice indicates to us that we are free, unbound by our supposed self-interests and natural inclinations. But this interpretation, as is, will not do, for at least two reasons. First, it doesn't seem that we need morality to show us that we can make such a free choice. Any pair of conflicting interests (should I have salad or soup with my dinner?) seems to be able to show us our freedom. But Kant insists that only the moral law can show to us our freedom; nothing else can. Second, this freedom has to be transcendental freedom for Kant, but this interpretation only seems to require a classical compatibilist conception of freedom, which only requires that we are free insofar as we are not constrained and can act in accordance with our desires and wishes (CPrR 5: 105). Transcendental freedom, on the other hand, requires much more; it requires independence from any causal chain of events that doesn't begin with the agent. The first problem, I will argue, can be answered by adopting an account of the fact of reason as an implicit awareness of something fairly specific, namely the Categorical Imperative, and by understanding why Kant thought that more ordinary examples of our apparent free choice will not do. But the second problem, about how it reveals to us our transcendental freedom is far more difficult, and though I may succeed in being able to show why Kant thought that our subjection to morality implied our transcendental freedom, I cannot reconcile this claim with his other claims that suggest that such freedom is impossible. That reconciliation might require appealing to transcendental idealism, which I cannot pretend to explain here because it is complicated and extends beyond the scope of this work.

I will argue, along with Rawls, that moral theory, as Kant conceives of it, has conditions of adequacy other than the one we reserve for empirical statements, because morality is essentially about what reason tells us to do. It is not about reason guiding us toward an independently given, objective moral reality. Once we understand Kant's project as trying

to find, within our capacity of practical reason, something that we ought to follow, we have all the proof we need to accept morality in our lives. Kant's evidence for the moral law lies in the fact of reason, which asserts that we take the Categorical Imperative to have a decisive influence on what kinds of things we choose to do. But we need to understand this fact, more as Allison does, as an *implicit* awareness of the Categorical Imperative rather than as an explicit recognition of it as the moral law—as if people all explicitly try to follow the formula of universal law—in order for Kant's argument to be plausible. Or so I shall argue. The first step is to explain what it means to act from the motive of duty.

4.2 Acting from duty

It is widely accepted that Kant held that an action has moral worth if and only if it is performed from the motive of duty.[4] Kant's famous examples in the first chapter of the *Groundwork* suggest both that if an action isn't done from duty, it doesn't have moral value and if it is done from duty, then it does (G 397-99).

What seems to be contentious among commentators is Kant's further claim that one must act *solely* from the motive of duty in order for the action to have moral worth. But what appears to be rather uncontroversial is what it means to act from the motive of duty, and so very few of them make explicit what they take it to mean. I think, however, that what Kant means by it is far from clear.

The obvious candidate is this:

> S performs an action A from the motive of duty if and only if S performs A from her belief that A is right.

But there is a textual reason to avoid making this interpretation. When Kant introduces his examples to distinguish acting from the motive of duty versus acting from other motives,

[4] See Aune (1979, 9) and Stratton-Lake (2000, 11).

he writes,

> I here pass over all actions that are already recognized as contrary to duty, even though they may be useful for this or that purpose; for in their case the question whether they might have been done *from duty* never arises, since they even conflict with it. (G 397)

Kant says that he doesn't consider examples of immoral actions, because he doesn't think it is possible to perform an action from the motive of duty if it is in fact wrong. But it is possible for a person to believe that an action is right when it is in fact wrong, and it is possible for a person to act on the belief that a particular action is right when it is in fact wrong. An action done from the belief that it is right cannot, for Kant, be sufficient for it to have been from duty.[5]

Given the consideration above, the following account of what it means to act from the motive of duty suggests itself.

> S performs an action A from the motive of duty if and only if (i) S performs A from her belief that A is right, and (ii) A is right.

But this revised definition will also not do. It is possible that S could correctly believe that A is right for the wrong reasons. Consider the following example, which is adapted from an example of Nomy Arpaly's (2002, 227). Oliver is a chess club extremist, and will do most anything, including murder, for his chess club. But Oliver believes that murder is always and only prohibited when it involves fellow club members. He wants to murder Stan for his overly defensive style of play, but he refrains from doing so on the grounds that Stan is a fellow chess club member. Here Oliver commits the right act (or refrains from committing

[5]One might think that there is another way of reading what Kant means in the passage above. He might mean that it is impossible to perform an action from the motive of duty if you *believe* it to be contrary to duty. On such an interpretation, we would be able to maintain the belief-only understanding of the motive of duty, but only at the high cost of committing Kant to the claim that Adolf Eichmann's actions, for instance, could be thought of as having moral value. We have a choice. We can choose to disagree with Kant and deny the close connection between acting from the motive of duty and moral worth. Or we can choose to maintain the connection and reinterpret what it means to act from the motive of duty. The first choice leaves Kantian interpretation and much of Kantian ethics behind. I will thus pursue the second option.

the wrong act) on his true belief that he ought not to murder Stan. But we are unwilling to think of his action as having moral worth. And thus I don't think we should think it was performed from the motive of duty either.

Oliver only happens to perform the right action. Even though Oliver's *belief* that it was right did move him to act rightly, he wasn't motivated by the right reasons. His belief that it was right only happened to be correct. But if acting on the belief that it is right is not enough, it hardly seems that adding the condition that he get it right as well would be sufficient. What we want is a deep connection between what makes an action right and what causes a person to act in its accordance.

We can look to the causal definition of knowledge for inspiration.[6]

> S knows that p if and only if the fact p is causally connected in an appropriate way with S's believing p.[7]

This close connection between what makes a claim true and what leads a person to believe it seems to be what is missing in our chess example: the extremist fails to have a certain responsiveness to the right reasons. Whatever the merits this definition has for "knowledge", I think it provides a deep and useful parallel for how we are to understand acting from the motive of duty. Thus, I propose the following account of what it means to act from the motive of duty.

---

[6]One might think that the considerations I've been making should lead in a different direction. For instance, one might think that I should simply add another condition, namely that the agent be justified in having the belief he has. And indeed, I will argue that if this condition is satisfied, along with the first two (the action is right and the agent has acted from the belief that it is right), then one has acted from the motive of duty. But that is because, as I will contend, Kant views morality to be about what reason requires us to do. I will, however, argue that neither this condition of being justified in one's moral belief nor the condition that the agent believe the action to be right is necessary to have acted from the motive of duty. So for present purposes, I choose to consider the causal account of knowing.

[7]See Alvin Goldman (1967: 369), "A Causal Theory of Knowing." One might worry that "an appropriate way" smuggles in requirements related to the traditional conception of justification, but Goldman intends no such implication. The agent, as he sees it, is not responsible for ensuring the appropriateness of the causal relation.

By using the causal account of knowledge as inspiration, I do not thereby mean to endorse it.

> S performs an action A from the motive of duty if and only if what makes A right appropriately moves S to perform A.

This definition has only one condition on the right side, but it tacitly assumes at least one other—that the action is right. As for the interpretive question, I think it must be the case that acting from the motive of duty requires the tacit assumption that the act be right because Kant says as much when he denies the possibility of acting from the motive of duty when the action being considered is in fact wrong.

But there is something noticeably absent in my definition: I have left out the condition that the agent S believes that A is right.[8] In fact, I think there are good reasons to leave it out. The implication is that it doesn't rule out the possibility that one might act from the motive of duty without knowingly believing that one's act is morally right. One might think that this is problematic as an interpretation of Kant's view, and perhaps problematic as an account of moral value, provided that acting from the motive of duty is a guide to moral value at all. To this worry, I reply that my definition does have the implication that one might be able to act from the motive of duty without knowing it—that one could perform an action from the motive of duty while not necessarily knowingly believing that the action is right. (By saying that the agent need not believe that the action is right, I intend all of the following to be possible: the agent could believe that the action is amoral, she could believe it to be wrong, or she may not have beliefs about its moral status at all.) But I believe that this is a virtue rather than a defect of my theory.

One apparent problem of my view is that it prima facie fails to respect the supposed value of acting from the motive of duty. In a now famous example, Barbara Herman writes,

> Suppose I see someone struggling, late at night, with a heavy burden at the back door of the Museum of Fine Arts. Because of my sympathetic temper I feel the immediate inclination to help him out. . . . [T]he class of actions that

[8] By leaving out this condition, I mean only to say that S need not *knowingly* believe that A is right. S may, nonetheless, *implicitly* believe that A is right.

> follows from the inclination to help others is not a subset of the class of right or dutiful actions. (4–5)

By acting from sympathy, we might choose to help the struggling person and thus perform the wrong action. On the other hand, if the motive of duty is present in our mind, then we would refrain from helping this person. Acting from the motive of duty thus better tracks our moral duties than the motive of sympathy.

But Herman understands our motive of duty as involving an express interest in the rightness of an action. She writes, "for a motive to be a moral motive, it must provide the agent with an interest in the general rightness of his actions" (6), and "for an action to *have* moral worth, moral considerations must determine how the agent conceives of his action (he understands his action be what morality requires), and this conception of his action must then determine what he does" (16). What this seems to mean is that acting from the motive of duty is acting from the belief that the action is right. But given the examples I've considered, I think we can now see why my understanding of what it means to act from the motive of duty is better suited to tracking our moral obligations than Herman's understanding. Acting from the belief that it is right is not nearly as good at tracking what is right and wrong as Herman seems to think. Being moved by the right-makers of an action better tracks whether an action is right or not. My interpretation, after all, *requires* the act to be the right act, whereas Herman's understanding only requires the agent to *believe* that it is the right act—and one's beliefs about what is morally required can go horribly wrong.[9]

[9]Herman isn't exactly clear whether acting from the motive of duty already requires the action be right. She does write that "an action has moral worth if it is required by duty and has as its primary motive the motive of duty" (16). But it is not clear whether she intends both of these conditions to be distinct and also necessary, or whether she thinks acting from the motive of duty requires that so acting is in fact dutiful. Let us say this. If on Herman's view, acting from the motive of duty is impossible without the action being right, her view still doesn't quite suffice. My example about Oliver, the chess club extremist, shows that doing a right action from the belief that it is right doesn't suffice to count as having acted from the motive of duty. If, on the other hand, acting from the motive of duty doesn't rule out acting from a mistaken belief that an action is right when it is in fact wrong, then my objection here goes through. What we still need is the connection between what makes an action right and what moves a person to act.

Before I consider other virtues of my account, we should examine what my account implies. What does it mean to say that what makes an action right is what moves the person to act? It implies that one might act from the motive of duty without knowing that she is acting from the motive of duty. It also implies that moral reasons, or whatever it is that makes an action permissible or impermissible, can affect what we do and how we act without being aware of it. Both of these implications are essentially neutral between competing normative ethical theories. Kant, of course, believes than an action is morally permissible if and only if the maxim underlying a proposed action passes the test of the Categorical Imperative (or CI-procedure). Suppose you are considering whether to perform a particular action for a particular reason. But somewhere in the back of your mind, you test your maxim (your reason for acting) against the CI-procedure, and it fails. If the fact that it fails holds you back from committing that action, despite what you knowingly express as your beliefs, then you have acted from the motive of duty. You do not need to be aware of what moved you to act in order to have acted from the motive of duty.

This may seem odd, and I suspect that some of the reluctance lies in a suspicion of what it means for thought processes to be in the 'back of one's mind'. I don't pretend to explain how that works or how it could work, but I will argue that we don't usually find it odd. We are more than willing to accept its parallel in cases of knowledge.

There are many things we justifiably believe without necessarily knowing why we are justified in believing them. For instance, I am justified in believing that I am now sitting at my desk, but I'm not sure what my justification for that belief is. And I'm not sure why I'm justified in believing it either. And to go further, we are sometimes wrong about the justifications. For instance, Descartes held that he was justified in trusting some of his sense perceptions (the clear and distinct ones) because an all perfect God would not deceive us so thoroughly (*Meditations* 37–41). I'm fairly certain that his justification is faulty, but I'm also fairly certain that he was justified in trusting his senses and the beliefs they gave him.

He was justified in believing he was sitting by a fire, that there was a tree in the distance, and so on, even though he didn't know why he was justified. That is, Descartes was justified in many of his sense perceptions while being wrong about what the justification for his beliefs actually are.

We are, I think, moved to believe certain propositions by the factors that make those propositions true, even if we don't know how they do so. I am justified in thinking that there is a desk before me, that I am now awake, and so on. The fact that there is a desk before me, in part, causes me to believe that there is a desk before me. The fact that the surface of the desk now reflects light also plays a part in causing me to believe that there is a desk before me. There are, of course, many facts which causally influence me to believe that there is a desk here.[10] And I may indeed be unaware of many of these facts. The point is, what makes it true that there is a desk before me plays a crucial role in causing me to believe that there is a desk before me, and that fact makes me justified in believing that there is a desk.

But the distinction between being justified in believing P and knowing what our justification for believing P is is not confined only to our sense perceptions. We could arrive at the belief that there are an infinity of prime numbers via Euclid's famous proof by contradiction, but then forgetting how to prove it, we construct an invalid proof which we mistakenly take to be valid. I submit that we would still be justified in believing that there are an infinity of prime numbers even if we are wrong about its justification. I suspect that this kind of discrepancy between what we believe and what we believe ourselves to believe is rather common, and a large number of epistemologists think so too.[11]

[10] I am here ignoring various issues about the use of the word "fact". The way I am understanding the word rests on how it is used in the correspondence theory of truth, which asserts that a proposition P is true if and only if it corresponds to a fact. In other words, the fact is what makes the proposition true, and it can have causal influence if it is a physical fact.

[11] Many epistemologists reject what is sometimes called the KK thesis, which states that for any proposition, if one knows that p, then one knows that one knows that p. For instance, Timothy Williamson (2000) rejects it. Of course, many epistemologists accept it as well; see Jaakko Hintikka (1970).

I think we can see this discrepancy in moral motivation as well. I think we can be moved by the reasons that make an action right even though we might not knowingly believe the action to be right. Arpaly, in her paper "Moral Worth", argues, convincingly I think, that it is common for people to act for the right moral reasons without it being the case that they are knowingly acting out of moral principles. She makes this possibility plain in her analysis of a story in Mark Twain's *The Adventures of Huckleberry Finn*. In the story, Huck Finn and Jim become friends, but Huck soon comes to the conclusion that he should turn his new friend in. Jim, after all, is a slave and 'belongs' to Miss Watson. For Huck, helping a slave escape is tantamount to stealing, which he believes to be wrong, an inviolable moral constraint. Huck resolves to turn Jim in, but when he has the opportunity to do so, he cannot bring himself to do so. Huck concludes that he is weak of will and unable to do what he takes to be right. But for us readers, we know that Huck does the right thing, and we are pleased that he could not bring himself to turn Jim in.

One way of understanding Huck's situation is to take him at his word (or at least at the words he tells himself). That is, Huck Finn is weak of will, and as such, he lacks genuine moral character. He has let his feelings of sympathy guide his behaviour and take precedence over what he takes to be moral and obligatory. On this understanding, Huck is a morally weak child who only happens to do the right thing. But according to Arpaly, whether his action is morally praiseworthy or not depends on how we interpret his motives, because there is another way to understand his action—as a response to moral reasons and thus as praiseworthy. She writes,

> Talking to Jim and interacting with him, Huckleberry constantly perceives data (never deliberated upon) that amount to the impression that Jim is a full person, just like Huckleberry himself. While he never deliberates on his perceptions, they prompt him increasingly to act toward Jim as a friend. Twain makes it very easy for the boy to perceive that Jim is very similar to him: Jim shares Huckleberry's language, knowledge, ignorance, and superstitions; and all-in-all it does not take the genius of Mill to see that there is no particular reason

> to think of one of them as inferior to the other. That Huckleberry begins to perceive Jim as a fellow human being is suggested when Huckleberry finds himself, to his surprise, apologizing to Jim—an action unthinkable in a society treating black men as subhuman.. . . He does not have the belief that what he does is right anywhere in his head—this moral insight is exactly what eludes him. Yet when the opportunity comes to turn Jim in, and Huckleberry experiences a strong reluctance to do so, his reluctance is to a large extent the result of the fact that he has come to perceive Jim as a person, even if his conscious mind has not yet come to reflective awareness of this. To the extent that Huckleberry is reluctant to turn Jim in because of Jim's personhood, he is acting for morally significant reasons. This is so even though Huckleberry knows neither that these are the right reasons nor that that he is acting from them. On this reading, he is not a bad boy who has accidentally done something good, but a good boy with imperfect knowledge. (2002, 229–230)

On this account, it is because Huck begins to see Jim as a person, and as an equal, that makes it difficult for Huck to turn Jim in. And this is so even though Huck doesn't realize how his views of Jim have changed. Thus, on this interpretation, Huck fails to turn Jim in not because he is weak, but because moral reasons move him to. Huck happens to be mistaken about what his motives are.

If this is possible, and I think it is, then it is possible to perform an action for the reasons that make it obligatory without being aware of the fact that your action is obligatory or even permissible. And like Arpaly, I think that this is more common than many moral philosophers seem to think. I think we are often wrong about what our own motives are, and nothing precludes the possibility that our motives could be moral without our knowing that. For instance, we all know people who perform kind and considerate things for others, but who at the same time, defend their actions as being ultimately motivated by self-interest. But instead of thinking that they are selfish or bad people, we tend to think that they are good people who are wrong about their own motivations. It would be hard to say that these people's actions have no moral value. That is one reason to leave out the requirement that one needs to act from the belief that one's action is right in order to act from the motive of

duty. Sometimes people act for moral reasons without being aware of it.[12]

Here's another reason. Suppose someone performs an action A for a reason R, and that R gives the conclusion that A is morally obligatory. I submit that the conclusion that A is morally obligatory should not be thought of as a reason for doing A. The reasons for doing A and the reasons why A is morally obligatory should be one and the same. As Philip Stratton-Lake puts it, "If an action ought to be done, then the reasons for doing it are the reasons why it ought to be done, and the fact that it ought to be done cannot be a reason why it ought to be done" (1). To put it another way, that an action is morally obligatory provides no reason to perform that action over and above the reasons that make that action morally obligatory. To think that it would is a little odd. This point is likely controversial, but it seems to me obviously true.

[12]This line of reasoning might invite the following question: Does Huck Finn believe that helping Jim was morally right? In explaining how Huck Finn's action could have moral worth despite his own professed believing that it was wrong, I relied on the plausible claim that Huck Finn is simply bad at understanding what morality requires and about what he takes himself to be doing. What we know is this: that Huck would not say "I did what I thought was morally right." But given my analysis and Arpaly's analysis of what goes wrong, it's less clear whether Huck thought his action was morally right.

We might say that Huck thought that his action was morally right, but that Huck did not *believe* that he thought it was morally right. In other words, we might say that Huck believed that A was right, but Huck did not believe that he believed that A was right. (That is, in Huck's mind, he holds P, but not $B(P)$.) This option is open to me, because all my argument requires is that Huck does not believe that he acted from a moral belief. (That is, all my argument needs is that Huck does not have in his head $B(P)$.) Pursuing this option requires denying something like a "belief" version of the KK-thesis, which would assert that S believes that P implies that S believes that she believes that P.

On the other hand, we might say that Huck did *not* believe his action was morally right. We could instead rely on the idea that the propositional attitude of believing is referentially opaque even when it comes to what is morally required. Classic examples of referential opacity go as follows: John believes that Muhammed Ali is the greatest boxer ever; Ali is Cassius Clay; therefore, John believes that Clay is the greatest boxer ever. This conclusion doesn't follow, because John might not believe that Ali is Clay and believing produces referentially opaque contexts. Here in Huck Finn's case, something similar might be going on. Huck Finn's action A is the only morally permissible action in his circumstances, but Huck would deny it. Huck Finn performed the act for the reasons that make A right, but, because of referential opacity, we might nonetheless say that Huck did not act from his belief that A is right, and Huck did not believe that his act was right.

So did Huck Finn believe that helping Jim was right, but didn't know that he believed it? Or did Huck Finn not believe that helping Jim was right? I don't know how to answer these questions, because I'm not sure which is the right analysis. Both of these analyses need further development. Nonetheless, I have argued that the following claims can be true simultaneously: (i) an agent could be motivated by the reasons that make an action obligatory, and (ii) an agent need not know that he acted from the belief that his action was obligatory. And that is all I need for my arguments here to work.

Let us compare our reasons for acting with our reasons for believing. Suppose you believe a proposition P for a reason Q, and that Q is a reason for thinking that P is true. Again, the fact that P, or the fact that P is true, should not form part of your reason for believing that P. The reasons for believing that P and the reasons why P is true should be the same. To put it another way, that a proposition is true provides no reason over and above the reasons for believing it. I think that this is fairly intuitive. There can, of course, be reasons to believe something that may have little to do with whether it is true. For instance, the argument in Pascal's wager does not try to persuade you to believe in God on the grounds that it is likely true that God exists. But if you are being persuaded to believe in the *truth* of a claim, the truth of the claim does not serve as an additional reason to believe it over and above the reasons that show such a claim to be true. The same follows about moral reasons. That an action is obligatory is not a reason to do it over and above the reasons that show that it's obligatory. Thus, the condition that one has to believe that one's chosen action is obligatory (or permissible) in order to be acting from the motive of duty is pointless. It adds nothing to the moral value of the action.

We might want to maintain that there is something important in saying "You ought to do it because it's the right thing to do." But it is hard to get at what that is. We often speak loosely in the analogous cases about belief and truth. We might say of a claim that "you should believe it because it is true". Sometimes what we mean by this is that we should value truth over our other interests, say if our pragmatic interests favour not believing in a truth. And that consideration might lead us to think that its moral parallel says that we should value our moral interests greater than our prudential or sympathetic ones. But I think we rarely mean this in the epistemic case, because people never deny that claim. Rarely do we hear someone assert, of a proposition, "It's true, but I don't believe it". (Examples of philosophers talking about Moore's paradox don't count.) If someone were to say such a thing, we would likely take him to be saying something incoherent, rather than making a

remark about how truth isn't the deciding factor on whether she should believe a claim or not. Rather, when we say that "you should believe it because it is true" we usually intend to reassert ourselves—to emphasize how high our degree of belief in our claim is. We usually make such an utterance when someone disagrees with us, and we take them to be wrong. We may say of an action "you should do it because it is the right thing to do", but this doesn't mean that it being right counts as a reason to do it. The reasons that show us an action to be right should move us to act on them. So when Kant says that we should do something because it is our duty, this means no more than to say that we should do something because we should do it.

There is, I grant, a certain force in making these kinds of claims. Suppose you are on vacation in Spain, and you've promised a friend to buy him a particular belt, made and sold by a local craftsman. But it turns out that finding the shop and getting to it proves to be harder than you originally anticipated. Nonetheless, you value your friendship and the promise you made, and so you make the extra effort to get the belt. Another friend, surprised by the effort you have exerted to fulfill a promise, asks you why you put so much effort into fulfilling your promise. And you reply "because I promised to." This answer works to some extent, even though your interrogator already knows this fact. It has some force; it goes some way to answering the query. But it doesn't go all the way, because part of the full answer to the query must involve explaining why it is the right thing to do. (To see this, we need only ask about situations where the action in question is controversial. Suppose we ask John why he punched Al in the face. His response "because it's the right thing to do" will not suffice.) But if the answer has force, we have to explain its force. And if it is forceful, then it seems that the claim "I do it because it's right" cannot be redundant. One obvious answer is that we have an interest in the rightness of an action, and that gives us a further reason to do it. But given that I don't think that knowing an action is right provides a reason to do it on top of the reasons that show it to be right, I must offer an

alternative explanation of the claim's force. Instead, I say that the claim "I do it because it's right" is indeed redundant, but it is its very redundancy which gives it force. When we say "I do it because it's right", we imply that no further reason is needed—that whatever reason shows an action to be right suffices as reason to do it. We are not providing a further reason to do the action, so much as pointing out that our reason suffices. We are asserting something *about* our reason for doing it. We are asserting that once a reason shows an action be right, all other considerations are irrelevant to the question of whether we should do it: we do it because it's right.

To recap: this is my account of what it means to act from the motive of duty.

> S performs an action A from the motive of duty if and only if what makes A right moves S to perform A.

The greatest difficulty for my interpretation is this: my interpretation simply sounds wrong. The phrase "acting from the motive of duty" just *sounds* as if it means that one is knowingly acting from the belief that the chosen action is obligatory (or permissible). But I argued above that if this is how we read what it means to act from the motive of duty, it fails to respect Kant's claim that immoral actions could ever be done from the motive of duty and it makes implausible the claim that only acts performed from duty could have moral worth. That said, I grant that the usual interpretation sounds better than mine.

One consideration that might mitigate against this worry is that Kant is sometimes translated as being concerned with "acting *from* duty" or "acting *out of* duty", which differs greatly from the phrase "acting *from the motive of* duty". It's hard to understand the English phrase "acting from the motive of duty" as "acting for the reasons that make that action right"; whereas it is easier to understand "acting from duty" in this way. Another consideration is that Kant is trying to distinguish different motives: that of sympathy, that of prudence, that of duty, and so on. Nothing about the phrase "the motive of duty" requires a conscious attempt to act in its accordance.

I concede that the examples Kant uses in the *Groundwork* to try to show the value of acting from the motive of duty have often seemed to readers to be pointing to the difference between what we do unconsciously and what we choose to do consciously after moral deliberation. And on this reading, Kant places moral value only on what we do consciously; whatever we do unconsciously or unthinkingly has no merit. I don't think this is the correct reading, but it's not to hard to read Kant as suggesting it. In one of Kant's examples, he begins by noting that we all have an "immediate inclination" to preserve our lives. Kant writes,

> . . . if adversity and hopeless grief have quite taken away the taste for life; if an unfortunate man, strong in soul and more indignant about his fate than despondent or dejected, wishes for death and yet preserves his life without loving it, not from inclination or fear but from duty, then his maxim has moral content. (G 398)

We might be led to think that the contrast Kant is drawing is one between the immediacy of an inclination and the conscious decision that we are making. But a little reflection quickly shows that this distinction is not sufficient for accounting for the difference between acting merely in accordance with duty and acting from duty. After all, Kant's first example is that of the prudent shopkeeper who decides to charge a fair price, because it is in his own best interest. This action could very well be the product of deep deliberation about how to make the most profit in the long term, yet Kant clearly thinks that so acting has no moral value. This should not be surprising: acting deliberately is of course no guarantee of acting morally.

But the point I've been making is this: knowingly acting from the belief that your action is right is insufficient and unnecessary for the action to have moral worth. I admit that Kant's examples of moral actions tend to involve a self-acknowledged decision to act morally. Much of the reason for that is that Kant's examples involve an internal conflict between two choices. It is much easier and much more natural to think of internal conflicts

as one between one's immediate inclination (or one's sympathy or prudence) and one's moral obligation.

And I suspect that another reason for reading Kant this way is a result of our own ethical views. Many of us modern readers today tend to disagree with Kant on our specific duties. For instance, we are less confident than he is that suicide is wrong. But given that we want to preserve a meaningful thesis in Kant's example, we are tempted to say that what matters in the example is not that suicide is wrong. Instead, we say that claim is irrelevant, and we say that the unfortunate man *consciously* chooses to do what he thinks is right, and choosing what you think is right is what has moral value. But reading it this way, as tempting as it is, precludes us from drawing the correct interpretation. Instead, the fact of an action's being right or wrong moves us to act in accordance with it or against it. All of his examples could be read, I contend, in my way. Or at the very least, my contrast between acting for the reasons that make the action right and acting for other reasons is preserved in all his examples and consistent with the lessons he draws. So my claim is that in his examples, Kant is not trying to draw attention to the conscious deliberative aspect of a moral decision, but rather to the motive that underlies our actions. Our moral actions require a moral motive for Kant, but they do not require those actions to have been done with knowing what those motives are.

Before I go any further, I wish to make something clear. In this section of the chapter, I have been arguing for two different claims. One is that Kant himself had the view I am attributing to him. And the other is that the view I'm attributing to him—that acting from duty means acting from moral reasons, which doesn't itself require knowing that one is acting from a moral belief—is an interesting and promising thesis in moral philosophy. I am, of course, trying to do both. But if it came to choosing between the two, I would maintain the second claim, and I would further say that this view is Kant*ian*, even if it's not Kant's. I could very well be wrong about what Kant, the man himself, actually believed.

But the view I am offering respects many of the important points Kant makes, and I believe that it reflects, at the very least, a strand in his thought.

That being said, Kant makes a claim in his *Groundwork* which poses a problem for my interpretation. Kant writes,

> . . . *duty is the necessity of an action from respect for law.* . . . Now, an action done from duty is to put aside entirely the influence of inclination and with it every object of the will; hence there is left for the will nothing that could determine it except objectively the *law* and subjectively *pure respect* for this practical law, and so the maxim of complying with such a law even if it infringes upon all my inclinations. . . .
>
> nothing other than the *representation of the law* in itself. . . can constitute the preeminent good we call moral, which is already present in the person himself who acts in accordance with this representation and need not wait upon the effect of his action. (G 400–1)

This passage seems to suggest that acting from the motive of duty requires being consciously aware that one is choosing to act in accordance with morality—that seems to be the meaning of the phrase "respect for law". The word "respect" tends to evoke a kind of acknowledged recognition and due regard. When we say "a respected academic", we mean someone who is well-regarded by others and known to be so regarded. Kant also contrasts between "objective" and "subjective". The use of the word "objective" here, referring to the practical law itself, seems to indicate its independence from us, while the use of the word "subjective"' seems to indicate an attitude to the moral law.

But I think we often respect certain theories, practices, people without our necessarily being aware of it. I would contend that most of us in the Western world have a deep respect for science, even if some of us do not believe it. It is not uncommon for many people to rail against Western medicine, for instance, claiming that it fails to acknowledge the wisdom of an ancient medical tradition. But often these same people will abandon the ancient herbs once the danger becomes more pronounced. For example, they may avoid doctors when they have mild illnesses, such as colds or rashes, but they do not hesitate to rely on those

same doctors when it comes to heart attacks or cancer.[13] There is a sense of the word "respect" which we often use to describe an unshakeable or foundational aspect to the way we live our lives. We respect the logic of an argument every time we are persuaded by it or feel compelled to argue against it, even if we claim to be suspicious of logic. We respect science every time we check the weather channel or open our refrigerator. Whether Kant intends to use the word "respect" as involving a conscious acknowledgement or not, he clearly intends this kind of respect for the moral law—the kind that underlies our behaviour. A being with a good will would always exhibit at least this kind of respect toward the moral law.

I want to return now to a point I made earlier in this chapter. Despite how the phrase "acting from the motive of duty" may sound, I have argued that we ought to understand acting from the motive of duty as necessitating being moved by what makes the action right. And I also argued that acting from the motive of duty need not involve knowingly acting in accordance with what one takes to be right or wrong. I rejected an account of acting from the motive of duty as insufficient, but rather than supplanting the insufficient account with further conditions, I advocated a wholly alternative account. More specifically, I rejected the account "S performs an action A from the motive of duty if and only if (i) S performs A from her belief that A is right, and (ii) A is right", and being inspired by the causal theory of knowledge in epistemology, I moved on to advocate "S performs an action A from the motive of duty if and only if what makes A right moves S to perform A." But in so doing, I skipped the more obvious inspiration of appealing to the traditional definition of knowledge. Here I wish to return to this idea because it is rich and complicated, and there is much to be learned by comparing it with my 'causal' account, especially with respect to the nature of Kant's ethics.

[13] What I say here doesn't imply that there is no value to any ancient medical tradition. Nor does it imply that there is no rational way to use both. But I think it's clear that there are people who claim not to trust science, while their actions show otherwise.

The traditional definition of knowledge is this:

> S knows that p if and only if (i) S believes that p, (ii) p is true, and (iii) S is justified in believing that p.

The moral parallel would be this.

> S performs an action A from the motive of duty if and only if (i) S performs A from her belief that A is right, (ii) A is right, and (iii) S is justified in believing that A is right.

This parallel will not do. I had argued above, while discussing Arpaly's Huck Finn example, that it was unnecessary for an agent to believe that her action is right in order for the agent to have acted from the motive of duty. But I will argue that these three conditions suffice to have acted from duty. That is,

> T: If (i) S performs A from her belief that A is right, (ii) A is right, and (iii) S is justified in believing that A is right, then S performs an action A from the motive of duty.

Even though I have so far emphasized the possibility that one can act from duty without knowing it, we must remember that it is just as possible to act from duty *knowingly*. I think T describes how this might happen. But we may worry about whether T is true, after all its epistemic parallel is open to Gettier counterexamples. Gettier counterexamples attempt to show that the three conditions for knowledge are insufficient. So let us look at one of Gettier's counterexamples for the traditional definition of knowledge, and then I will try to construct a parallel counterexample. First, Gettier considers the case of Smith who is applying for a job. Gettier stipulates that Smith is justified in believing that Jones, the other candidate, will get the job. Smith is also justified in believing that Jones has ten coins in his pocket. Smith thus infers and believes that the man who has ten coins in his pocket will get the job. It turns out that Smith will get the job, and that he has ten coins in his pocket. So Smith believes that the man who has ten coins in his pocket will get the job, his belief

is true, and it is justified (on the seemingly benign assumption that the logical inference of two justified beliefs is also justified). But we would hardly say that Smith *knew* that the man who has ten coins in his pocket will get the job. Thus, the three conditions of knowledge do not suffice.

One compelling diagnosis of the situation is that there is a discrepancy, a lack of any real connection, between what makes it true that the man with ten coins will get the job and Smith's belief that the man with ten coins will get the job. The causal account of knowledge thus requires a deep connection between what makes a claim true and why the person believes it in order for there to be knowledge. This consideration is compelling on its own, and motivates my account of acting from the motive of duty. The Gettier counterexample seems to make this point clear.

But how would the moral parallel work? Suppose Brown is considering hiring one of two candidates, Smith or Jones. Brown is justified in believing that Jones is the more qualified of two candidates, and he is also justified in thinking that Jones is the one with the blue hat. Brown is thus justified in believing that the man in the blue hat is more qualified. So he hires the man in the blue hat, but unbeknownst to him, it is Smith who is in the blue hat and it is also Smith who is more qualified. It appears that Brown acted on his belief that the man in the blue hat was more qualified, the man in the blue hat was more qualified, and Brown was justified in his belief. We might agree with Gettier that Brown did not *know* that the man in the blue hat was more qualified. But let us ask the moral question, "Did Brown act from the motive of duty?"

I think we want to say yes. If we ask "Does Brown's act of hiring the man in the blue hat have moral worth?" I think we want to answer yes again. (All this assumes that it is morally required for Brown to hire the one who is most qualified, and that Brown acted for that moral reason. He may very well have self-interested reasons to do the same thing.) The parallel counterexample seems to fail. We want to say that if Brown is genuinely justified

in his belief in his action then his action is indeed justified. He happened to perform the right act, and it was relatively unconnected from his reasons for thinking that it was the right act. There was something lucky in Brown's action. But still, we want to say that if he was truly justified in his belief, his acting on it counts as having moral worth.

On the other hand, let us consider the historical examples of people who commit horrendous acts because they believe it to be right, such as Adolf Eichmann. We might grant that they believed their acts were right, but we do not think that they were justified in their beliefs. Somehow, we still say that they should have known better. Our example of the chess extremist does not have justified beliefs about morality, and that is part of what explains why his actions are wrong. He too should know better. They were unjustified in thinking what they thought. But in regular Gettier examples, we usually have little trouble accepting the stipulation that the believer was justified. But why? What is the difference?

Kant's theory helps explain the difference. For Kant, ethics is essentially about justification, and what we really have justification to do, whether we know it or not. And since justification, at least when it comes to ethics, is essentially internal, whether your action has moral worth or not, depends entirely on its justification and whether it passes certain standards. In this way, justification in ethics resembles justification in logic or math. No one, for instance, is justified in thinking that there are a finite number of prime numbers, because a person's basic beliefs about the natural numbers logically entails otherwise. To take a more straightforward example, no one, for instance, is justified in thinking that because it is true that A, it must also be true that not-A.[14] This might lead one to think that

[14]These points depend on whether one can be justified in believing a claim on the basis of testimony. For instance, suppose Ernie, a well-respected mathematician, tries to play a bad joke on Bert, an ordinary layman, by trying to convince him that there are a finite number of prime numbers. Suppose Bert, lacking confidence in his own mathematical ability, believes Ernie. It is arguable that Bert is justified in believing there are a finite number of prime numbers. Let us avoid this issue by distinguishing between two kinds of justification for belief. Justification-1 is justification that does not involve relying on the testimony of others. Justification-2 is the kind that *can* involve relying on the testimony of others. (Justification-1 is intended to be a proper subset of justification-2.) I believe that when we use the word "justification" in ordinary parlance, we mean "justification-2", because we usually grant that ordinary people are justified in accepting the theory of General

the fundamental principle of morality, in particular on Kant's view, is a logical truth. That doesn't follow, but what does seem to follow is that, for Kant, the fundamental principle of morality is a priori, or to put it another way, a requirement of reason.

Thus, the difference between my 'causal' account of acting from the motive of duty and the moral parallel of the traditional definition of knowledge is small, because what makes an action right is that it meets the ultimate justificatory standard, known as the Categorical Imperative. And the difference between the causal account of knowledge and the traditional definition of knowledge is significant, because the justification the knower has for believing a claim, especially an empirical fact, may have very little to do with what makes that claim true.

Understanding the Categorical Imperative as a requirement of reason also helps us reconcile the 'causal' account of acting from duty with the 'traditional' account in another way. I have argued that we can be moved by the right-makers of an action without our knowing it. I have also argued that on Kant's view, the essential principle that makes an action right or wrong, the Categorical Imperative, is a justificatory standard or a requirement of reason. We might worry whether these two claims are compatible, because we might think that our processes of reasoning are always transparent to us. Thinking of the Categorical Imperative as a requirement of reason actually helps us explain why morality can influence us without our knowing it, because requirements of reason can influence our beliefs, in an undetected way, just as external facts can.

Imagine a person who professes believing in a peculiar postmodernist theory that says that science is bunk, and that it is rational (or at least, not irrational) to believe whatever one's friends believe. Suppose further, that she wants to believe that the Earth is flat, because all her friends are Flat-Earthers. Try as she might, she cannot bring herself to believe

---

Relativity, for instance, because they are justified in trusting physicists. So to stipulate, in this context, what I have been meaning by the word "justification" is "justification-1"; it does not involve the testimony of others.

that the Earth is flat, because she is a former astronomy student, who has a long history studying the laws of physics, the evidence for the basic claims of astronomy, and what they entail. She may even say "I know that it's irrational to believe that the Earth is round and spherical, but I can't help it." I would argue that insofar as her belief is moved by the evidence, she is not, in fact, being irrational in her belief in a round, spherical Earth. (She may, however, be irrational about other things.) Her belief is rational; she is moved by the requirements of reason to reject the Flat-Earth thesis, even if she herself would not describe it as such.

We do not need to be consciously aware of the requirements of reason for them to have an influence on what we believe. Further, being consciously aware of their influence on our beliefs does not make them more or less rational. Whether one of our beliefs is rational or not depends on the reasons we believe it, and not on what we take our reasons for believing it to be. I will grant that consciously trying to be (theoretically) rational may lead one to be more (theoretically) rational, just as consciously trying to be moral may lead one to be more moral. But just as knowingly believing that P is rational does not add to the rationality of believing that P, knowingly believing that one's action A is morally right does not add to the moral worth of performing that action A.

Consider a revised Huck Finn example, where Huck Finn* comes to the realization, quite self-consciously, that helping Jim is morally required. He thinks to himself, "No, wait. Jim is a person like me. He has beliefs, desires, and wishes. He has feelings like I do. He is capable of making decisions, and he cares for others. I should regard him as a moral agent, as worthy of respect just like any other human being. It would thus be morally reprehensible to turn him in." Huck Finn* thus commits the same action as did the original Huck Finn. I submit that Huck Finn*'s action has no more moral worth than the original Huck Finn's action; this awareness adds nothing to the moral worth of his action. They are both moved to act by what makes that action right. I may grant, that over time, Huck Finn*

is more likely to get better at being moral, but there is no guarantee of it.

I have argued that acting from the motive of duty means acting for the reasons that make that action right, and since what makes an action right or wrong is, on Kant's view, the Categorical Imperative, we thus have the following, fuller, thesis of what it means to act from the motive of duty:

> S performs an action A from the motive of duty if and only if the Categorical Imperative has a decisive and appropriate effect on whether S performs A. (By 'appropriate', I mean that if it fails, S refrains from performing that action.)

Given that I've also argued that requirements of reason can influence us without our being aware of it, and that the Categorical Imperative is a requirement of reason on Kant's view, this means that an agent need not be aware that his action is right, though he could be, in order to have acted from the motive of duty.

4.3 On Hume's dilemma

I have so far argued that one need not know that one is acting from duty in order to be acting from duty. My contention is controversial, and one might think that by insisting on it, I neglect certain advantages of the more traditional account. You might ask, Isn't there some value in deliberately choosing what we take morality to require? I do not deny that there is value, but I do not think it is moral value exactly. Or rather, I think its value cannot be translated in any straightforward way to the value of a particular action. Self-consciously caring about morality has its value largely because it encourages care in one's moral thought. It helps develop one's moral perception because it makes certain features of our world more salient. By caring about morality, we might, for instance, become more attentive and feel more displeasure at cruel and insulting jokes. We might also notice more of the relevant features that affect the rightness or wrongness of our actions. We would

likely also become more motivated to act out of moral concern. All this I grant. These are all ways of honing our moral attention and encouraging our moral behaviour (though things can still go horribly wrong), but these facts remain consistent with my contention that the capacity to be moved by what makes an action right or wrong is what indicates good will. If one is moved to perform an action by what makes it right, then one's action has moral worth—and that doesn't require that the agent be aware of it.

You may grant all this, but you might think I have missed something. There are, after all, interpretations of Kant's argument which attempt to show the value of acting from the belief that one's act is right and the necessary role that acting from such a belief plays in Kant's theory. Most notably, Korsgaard (1996), in her paper "Kant's Analysis of Obligation: The Argument of Groundwork 1", relies on the assumption of this self-conscious adoption of morality to answer a problem that Hume posed in his *Treatise of Human Nature.* To understand her proposal, we need to understand the problem Hume posed, which Korsgaard calls Hume's dilemma. It begins by considering two apparently plausible claims:

(A) It is motives that essentially make an action right or wrong.
(B) The primary motive of virtuous action is the motive of duty.

We may come to accept (A) because we tend to judge whether a person's action was morally worthy depending on how we conceive of the person's motive. If he is trying to do good, then we tend to think his action good, and if he was trying to do bad, we think it bad. As Hume puts it,

> 'Tis evident, that when we praise any actions, we regard only the motives that produced them, and consider the actions as signs or indications of certain principles in the mind and temper. The external performance has no merit. We must look within to find the moral quality. This we cannot do directly; and therefore fix our attention on actions, as on external signs. But these actions are still considered as signs; and the ultimate object of our praise and approbation is the motive, that produc'd them. (T 477)

But we also tend to think that consciously doing our duty plays a crucial role in our moral behaviour, and thus (B).[15] Hume argued that these two claims cannot be accepted together.

If both (A) and (B) are true, then it seems to be impossible to know what one's duty is. Let us say that you are trying to determine what the right thing to do is. (A) tells you to be guided by the motives. There are certain motives which make action virtuous, and then you try to find out what a person with those virtuous motives would do. So you look for the right motives. (B) tells you that the primary motive of virtuous action is that you do it because it's right. We are in a circle. As Hume puts it, "to suppose, that the mere regard to the virtue of the action, may be the first motive, which produc'd the action, and render'd it virtuous, is to reason in a circle" (T 478). There seems to be no way to know what is the virtuous action. According to Korsgaard's understanding of Hume, you need to know "which actions are virtuous before you can do them with regard to their virtue or rightness" (1996, 48). Both (A) and (B) together imply the following:

(C) It is acting from the motive of duty which makes an action right.

According to Korsgaard, this threatens to be an empty formalism. But the consequences are worse than Hume or Korsgaard seem to think. The two claims, (A) and (B), imply that any action is right so long as you believe it is right and act from that belief. Ethical relativism follows. The dilemma is this: If (A), you can't get (B). And if (B), you can't get (A). You would like both, but you can't have both without absurd results.

According to Korsgaard, Hume and other sentimentalists reject (B) and choose (A). They thus say that it cannot be the motive of duty which is the primary motive. But since (A) says it is motives that essentially make actions right or wrong, we need another motive or another set of motives which are the primary motives of virtuous actions. For the

[15] I have given reasons to think that (B) isn't true, at least not as it is usually understood. But I have no qualms accepting that it plays a large role in our moral life, especially the large, not insignificant part of it where we are consciously trying to act morally.

sentimentalist, our way of figuring out which motives are virtuous involves a moral sense, by which we approve or disapprove of various motives. Korsgaard believes that the sentimentalist position suffers from two problems in addition to not being able to accommodate (B). The first is that on the sentimentalist view, no action can be necessarily right, because whether it is right or wrong depends on our moral sense and what it happens to approve or disapprove. If we had a different moral sense, different actions would be right or wrong. She claims that this makes right actions only contingently right. She writes, "So the action is not *necessarily* right. But then it is hard to see how it can be obligatory. An obligatory action is one that is binding—one that is necessary to do. But if the action is not necessarily right, how can it be necessary to do?" (1996, 48) Her rhetorical question ends her objection.

I don't think this objection works, because her rhetorical question can be answered. The sentimentalist can appeal to hypothetical or relative necessity, to be distinguished from absolute necessity. (This is an old distinction, which I discussed in my first chapter. Leibniz appealed to such a distinction.) The sentimentalist can assert that *given* the moral sense we have, it is necessary to perform the action. It is not absolutely necessary, but relatively necessary. And that is all requiredness of the action consists in. Korsgaard might argue that we need categorical or absolute necessity to have genuine moral obligations, but it is arguable that no action is categorically or absolutely necessary. After all, an action is only right or wrong given particular circumstances, which themselves may be contingent. For instance, helping people move furniture out of a house may be permissible, but that is *conditional* on the assumption that these people are not committing an act of robbery. Whether an action is right or wrong always depends on a set of contingent circumstances, and for the sentimentalist, those circumstances include the nature of our senses.

Korsgaard offers a different but related objection against the sentimentalist's choice, but it is couched again in rhetorical questions. She writes,

> If the moral motive is simply a natural affection such as benevolence, we can really have no obligations. For how can there be an obligation to have the motive which gives us obligations? And how can we be obliged to perform the actions, unless we are obliged to have the motive that produces them? (1996, 49)

This is how the argument seems to go.

> (A) It is motives that essentially make an action right. (The Sentimentalist's choice.)
>
> P1: If it is motives that essentially make an action right, then we cannot have an obligation to have those motives.
>
> P2: If we do not have an obligation to have the motives, then we do not have an obligation to perform the actions.
>
> Conclusion: We have no obligation to perform the actions.

The conclusion seems to follow from the two premises. But it is hard to see why she thinks P1 or P2 is true. P2 is subject to the same objection as above—that we can have relatively necessary obligations: we may not be obliged to have the motive, but given that we do have the motive, then it is necessary to perform the action it issues. Either way, she finds the sentimentalist's decision uncompelling.

Korsgaard also finds the rationalist's choice uncompelling. The rationalist chooses option (B)—that the primary motive of virtuous action is the motive of duty—but then he must reject (A) on the grounds that we wouldn't know what our duties were. The rationalist is thus compelled to deny that right actions are essentially defined in terms of the right motive. So for rationalists, what makes an action right or wrong must be intrinsic to the actions themselves, independently of the motive. This result is unacceptable to Korsgaard, because it implies an objectionable kind of realism. She writes,

> But now the rationalist is saddled with the view that rightness is (to speak anachronistically) a non-natural property, inherent in the actions, and intuited by reason. In this way, rationalism seems to entangle us in a metaphysical moral realism, as well as an epistemological intuitionism, which are both unpalatable. (1996, 52)

Many realists are perfectly willing to accept such results, but Korsgaard takes them to be unacceptable. Korsgaard doesn't mention it, but it is also unlikely that Kant would accept the rationalist account, because he would be sceptical of a faculty of intuition that wasn't essentially empirical.

Korsgaard accepts Hume's contention that (A) and (B) together imply (C), which says that it is acting from the motive of duty which makes an action right. But Korsgaard believes that this is not an empty formalism, and that Kant's analysis can give us obligations. She does this by trying to establish the following claims:

> (1) The good-willed person does the right thing because it is the right thing. (1996, 60)
>
> (2) The rightness of an action lies in the legal character of its maxim. (1996, 61)
>
> (3) The source of the legal character of a maxim must be intrinsic to the maxim. (1996, 62-63)
>
> (4) The legal character of the maxim just is whether it passes the CI-procedure. (The CI-procedure (and perhaps, the person's own will) is the only possible intrinsic source that can give a maxim its legal character.) (1996, 63)

Therefore, Korsgaard contends that the dilemma is answered because accepting (A) and (B) doesn't produce an empty formalism. Rather, accepting both (A) and (B) helps produce a rule with moral content, the CI-procedure, provided that we accept Kant's analysis of the concept of obligation. Here on Korsgaard's view, doing something because it's right means being concerned with the legal character of your maxim. But the legal character of a maxim, or of any claim at all, must be given intrinsically. For Korsgaard, any external force cannot give us any genuine normativity, because its force is always conditional on our accepting it as a force. Even if an external force carries with it a sanction, it remains conditional on our accepting the sanction as action-guiding. Since Korsgaard believes an action only has force once we accept its force, the normativity or "legal character" of a maxim must be given intrinsically. Given this instrinsic-ness requirement, the only way a maxim can have

legal character is by taking the form of a law, meaning that it is universalizable. Hence the CI-procedure. Thus if you know what acting from the motive of duty entails, then you know what your duties are, because the CI-procedure has content. Hence, you can accept both (A) and (B). Kant's view apparently avoids the dilemma.

Korsgaard's argument is appealing, and I think there is much that is right about it, especially her account of how Kant understands the concept of obligation in the *Groundwork*. (In my second chapter, I had already discussed (4) and expressed my scepticism about whether valuing the mere form of a law amounts to valuing the formula of universal law (72–4).) But Korsgaard's argument suffers from an ambiguity. When Korsgaard writes that "the good-willed person does the right thing because it is the right thing", it could be understood in two ways. For Korsgaard, it doesn't mean, as I might suggest, that the fact that makes it right moves the person to act. Rather, what Korsgaard means, is that the person acts from the belief that the action is right. For her, the agent is clearly conscious of what he is doing: the good-willed person does the right thing because he believes it to be right. But here, I think, is the nub. The person may not know what the rightness of an action implies. Korsgaard's argument moves from the agent's *belief* that an action is right to what *actually* makes an action right (or least what makes it right according to Kant).

This is a problem because the good-willed person may not know that (2)-(4) are true. He will remain stumped as to what he ought to do. Let us look at Korsgaard's argument for (3) to elaborate further. She argues that external sources cannot provide any genuine legal character to an action. She writes,

> Suppose . . . that the External Law is in force because it is the law of God's will. This is supposed to be what makes it normative for me. But how? God's will is only normative for me if it is the law of my own will to obey God's will. This is an old Hobbesian thought—that nothing can be a law for me unless I am bound to obey it, and nothing can bind me to obey it unless I have a motive for obeying it. But Kant goes a step further than Hobbes. *Nothing* except my own will can make a law normative for me. Even the imposition of a sanction

> cannot bypass my will, for a reward or punishment only binds my will if I will to get the reward or avoid the punishment—that is, if I make it my maxim that my interest or preservation should be a law to me. (1996, 65)

If the state orders me to pay taxes, then it only binds me insofar as I accept it as binding. If they threaten to punish me for failing to pay my taxes, it only binds me on the condition that I *choose* not to be punished. But it is hard to understand her claim "nothing except my own will can make a law normative for me". It seems to amount to the following: that I must *believe* an action to be obligatory in order for it to be obligatory at all. It's hard to read her otherwise. But this sentence has an absurd logical consequence. It implies that if I don't believe an action to be obligatory, then it isn't. We get moral relativism again.

The most obvious, and I think right, way to avoid this consequence is to say the following: my will makes laws normative for me without my necessarily knowing it. It isn't about what I deliberately and knowingly choose as my own law. Rather it is what my will regards as a law, independently of whether I know that or not. Hence, simply being a human being means having a will, and that automatically means that we accept (perhaps only tacitly) certain implications—say, if Korsgaard is right about (2)–(4), which assert that what makes an action right is whether its maxim passes the CI-procedure.

This is why we should abandon how Korsgaard understands what the good-willed person is doing. (It isn't necessary for her argument anyway.) Acting from the belief that an action is right doesn't ensure that it is what makes an action right which motivates the good-willed person's action. So the good-willed person, on her view, cannot automatically act rightly, because the good-willed person could easily be wrong about what the right actions are. What we need instead is to say that the good-willed person is moved by what makes actions right. That is how we should understand what it means when we say "the good-willed person does the right thing because it is right". We should not understand this as meaning that the good-willed person believes the action to be right, but the fact of its

rightness moves her to perform the right action.

As I understand Korsgaard's argument, it is essentially because the concept of obligation isn't empty for Kant that we have no problem accepting both horns of the dilemma. But Kant's analysis of the concept of obligation is independent of the two claims in the dilemma. So while I might accept Korsgaard's interpretation of Kant's analysis of the concept of obligation and how it might give rise to the Categorical Imperative, I think Kant would simply not endorse both claims of the dilemma.

The answer to the dilemma is easier than Korsgaard makes it out to be. I do not believe that for Kant what makes an action right or wrong is the motive, or at least not exactly. Korsgaard conflates Kant's distinction between *what makes an action dutiful* with *what gives it moral worth*. Kant endorses such a distinction when he says that *acting merely in accordance with duty* but not from duty has no moral worth. This would be nonsensical if there was no distinction between what gives an action moral worth and what makes it a duty. On my view, Kant would simply reject the first horn of the dilemma (and also the second one). What we have instead is this:

> (A') It is motives which determine whether an action has *moral worth*.
>
> (B') The primary motive of virtuous action is the motive of duty (which means, on my view, that what makes it right or wrong plays a crucial role in moving or constraining how the person acts)

These two claims do not produce any dilemma.

I have argued that if we are attentive to the distinction between what makes an action dutiful and what makes it have moral worth, as Kant maintains, then Hume's dilemma dissipates. But I have so far only tried to explain what it means to have performed an action from duty. If I am to maintain the distinction, I need to give a sketch of what makes an action dutiful. More specifically, I must be able to explain what happens when an agent performs an action in accordance with duty, but not from duty.

On the face of it, the question of what makes an action dutiful may appear straightforward. Kant provides the formula of universal law, which provides us with, following Rawls, what I've been calling the CI-procedure. As I wrote in the last chapter, the CI-procedure is a three step procedure that involves first determining the maxim that underlies the action you propose to take, second recasting that maxim as a universal law of nature governing all rational agents, and third considering whether you could or would rationally will to act on your original maxim in such a world. If not, then you have a duty to refrain from so acting. And if so, then your action is permissible. But even if all that is clear, it is less clear how we are to understand the possibility of an agent who performs a dutiful action, but not from duty.

Consider Kant's example of the prudent grocer. He makes it part of his policy to charge a fair price to inexperienced customers, and to refrain from taking advantage of them. But when we assess the prudent grocer's action and assert that it is dutiful, even if it is not done from duty, do we use the CI-procedure to test *his* maxim—the one that guides *his* action? We might want to answer yes. After all, the idea of a CI-procedure that tests our maxims assumes that some maxims can pass the CI-procedure and some can fail. So one could very well act on a permissible maxim without knowing that it is permissible (or that it passes the CI-procedure). The universalizability of an action may not come into the agent's head at all, but he has nonetheless happened to act on one that does pass the procedure. For instance, let us consider a maxim that Korsgaard considers permissible: "I will keep my weapon, because I want it for myself" (Korsgaard (1998), xv). Korsgaard uses this example to illustrate a different point, but it seems to me that it is possible that somebody could act on such a maxim without realizing that it is permissible.[16] One could intend to keep his

[16]She compares three maxims: the one above along with the following two.

(a) I will keep your weapon, because I want it for myself.
(b) I will keep your weapon, because you have gone mad and may hurt someone.

Her point is that the lawlike character of an action rests on the way the parts are combined, and that explains

weapon because he wants it for himself, without the issue of its universalizability affecting his choice. So it must be possible for a person, by mere chance, to act on a maxim that passes the CI-procedure. In other words, it must be possible for someone to perform an action from a permissible maxim, without either knowing that it is permissible or without the conditions of its permissibility affecting his choice. But how are we to describe such a person and his action? The only option seems to be that he has acted in accordance with duty but not from duty.

Alternatively we might want to claim that judging the prudent grocer's action does not require using the CI-procedure to test *the grocer's* maxim. We do something rather different. We ask if his action is in accordance with a maxim that we already accept to be permissible. This is O'Neill's suggestion (1989, 130-131). Much of the rest of her account is the same as mine. She accepts, as I do, that the formula of universal law is a test for the permissibility of a maxim. But to assess whether the grocer's action is done in accordance with duty, we do not assess *his* maxim but whether his action conforms with, or apparently exemplifies, *a* maxim that we know to be morally permissible. If it does, then it was done in accordance with duty. It is a further question whether the morally acceptable maxim guided the grocer's action. If so, then it was also done from duty; otherwise not. (It should be noted that O'Neill also stresses the possibility that the "maxim of an act may be a principle that embodies *no* description under which an agent consciously acts" (130). She is here concerned with the possibility that we could think we are acting from a moral maxim, when we are, in fact, not. Kant is suspicious of our capacity for self-knowledge, and we can never know which maxim actually guided our actions.[17] So due to our lack of

why (a) is a bad maxim, even though it contains the purpose of a good maxim (c), and the act of another good maxim (the one I used in the main text).

[17]In the *Religion*, Kant writes, "Assurance of this [the adoption of moral law as the ground of our actions] cannot of course be attained by the human being naturally, neither via immediate consciousness nor via the evidence of the life he has hitherto led, for the depths of his own heart (the subjective first ground of his maxims) are to him inscrutable" (6: 51).

self-knowledge and the possibility of error about what morality requires, she thinks it is possible we think we are acting from a moral maxim when we are in fact not. Though she doesn't mention it, her view also has the opposite implication as well: that it is possible to act from a moral maxim without knowing that one has acted from a moral maxim.)

Both my suggestion and O'Neill's account seem plausible to me. It seems to me that nothing in Kant's theory precludes the possibility that an agent would accidentally act on a maxim that is permissible. And it also seems possible to me that a person has performed an action that conforms with a morally permissible maxim, but for non-moral reasons. In both cases, I want to describe their actions as in conformity with duty, but neither were done from duty. But the two accounts do not have the same extension. In particular, it seems that the first description gives an extension which is a subset of the second's extension, because while any person who has happened to act on a maxim that can pass the CI-procedure has also performed an action that seems to exemplify a moral maxim, it is not the case that any person who has performed an action that seems to exemplify a moral maxim has acted from that maxim. The man who wants to keep his weapon for himself has acted from a morally acceptable maxim, but doesn't necessarily know it. The prudent grocer, on the other hand, may have acted from a non-moral maxim, but his action is in accordance with one. Still the two accounts give different positive accounts of what it means to have done a dutiful action.

Resolving this may amount to resolving, what some take to be, an essential project of understanding Kant's ethics: what is it that makes an action right or wrong? Fortunately, I don't need to resolve this question here. All I need is to suggest that both of the two most plausible accounts of the conditions under which an agent has acted in accordance with duty but not from duty are consistent with my account of what gives an action moral worth.

4.4 The Fact of Reason

I had argued above that to perform an action from duty is to act from moral reasons, which may or may not involve consciously being aware of acting for such reasons. And I had also said that the fact of reason is the fact that we can act from duty. And given Kant's account of what makes an action permissible or impermissible—whether its maxim passes the test of the Categorical Imperative—here's my account of the fact of reason.

> The justificatory standard of the CI-procedure can have a decisive and appropriate effect on whether a human being performs an action. (By "appropriate", I mean that if it fails, S refrains from performing that action.)[18]

In this section, I want to address three common concerns about the fact of reason. For one, Kant seems to claim that the fact of reason justifies morality.[19] I will try to make sense of how the fact of reason can do this. Second, I will try to pinpoint the difference between Kant's argument in the *Groundwork* and his argument in the second *Critique*. Third, I will try to make clear why Kant thinks that only the moral law could show us our transcendental freedom, and why nothing less than transcendental freedom will do.

Again, the most significant passages concerning the fact of reason are these:

> (1) Consciousness of this fundamental law may be called a fact of reason because one cannot reason it out from antecedent data of reason, for example,

[18] I take this to be roughly equivalent to the following claim:

> We (implicitly) respect the Categorical Imperative as having an authority over our actions.

These are roughly equivalent, because the Categorical Imperative's effect on our behaviour can be characterized as respect, because it's effect on us isn't decisive, though we might regard it as authoritative. It should be noted that my account resembles Rawls's account to some degree. He wrote that the fact of reason was "the fact that, as reasonable beings, we are conscious of the moral law as the supremely authoritative and regulative law and in our ordinary moral thought and judgment we recognize it as such" (260). But my account differs in two important respects: (1) it doesn't require that we know that the Categorical Imperative is the moral law or even that it has anything to do with morality, and (2) it doesn't require that we know that we are acting in accordance with the Categorical Imperative.

[19] More accurately, Kant says that "the moral law is. . . a fact of pure reason" and that "the moral law cannot be proved by any deduction, by any efforts of theoretical reason" and that the moral law "is firmly established of itself" (CPrR 5: 47). These claims suggest that the fact of reason serves to justify morality.

> from consciousness of freedom (since this is not antecedently given to us) and because it instead forces itself upon us of itself as a synthetic a priori proposition that is not based on any intuition, either pure or empirical... (5: 31)
>
> (2) This Analytic shows that pure reason can be practical—that is, can of itself, independently of anything empirical, determine the will—and it does so by a fact in which pure reason in us proves itself actually practical, namely autonomy in the principle of morality by which reason determines the will to deeds. At the same time it shows that this fact is inseparably connected with, and indeed identical with, consciousness of freedom of the will... (5: 42)
>
> (3) Moreover the moral law is given, as it were, as a fact of pure reason of which we are a priori conscious and which is apodictically certain, though it be granted that no example of exact observance of it can be found in moral experience. (5: 47)

These various passages are not exactly consistent with one another. They suggest that the fact of reason could be any of the following: (i) consciousness of the moral law, (ii) autonomy in the principle of morality, (iii) consciousness of freedom of the will, and (iv) the moral law itself. It will be hard to reconcile all of these without ascribing a certain amount of looseness to Kant's words.

The fact of reason is, I claim, the fact that the right-makers of an action *can* move us to act. They impinge on our consciousness and affect our behaviour. But I don't think my claim makes sufficient sense without understanding Kant's particular conception of morality. I believe that once we have an understanding of Kant's metaphysical view regarding ethics and accept my claim about the content of the fact of reason, we can reconcile the discrepancies about what he says regarding the fact of reason. I also think that in so doing we get a clearer picture of how the fact of reason is meant to justify our moral theory, but at the same time, we can understand the limits that Kant seems to think it can do. For instance, we can make sense of Kant's remark that "no example of exact observance" of an instance of the fact of reason can be found.

To give a brief sketch, I think Rawls and Allison are largely correct about the metaphysical aspects of Kant's view, and in particular how Kant's conditions of adequacy for a

moral theory differ from others—notably, how it differs from rationalists, sentimentalists and moral anti-realists.[20] I will differ from Rawls and reject his constructivism insofar as it offers a substantive metaphysical view. I will also reject Rawls's interpretation of the fact of reason as requiring our being aware of our consciousness of the moral law, favouring instead only the idea of being conscious of the moral law, without our necessarily being aware of it.

A common way of conceiving of the adequacy of a moral theory is to conceive of it in much the same way we conceive of the adequacy of an empirical claim or scientific theory. Let us say that this means that there are facts, objectively and independently of our consciousness, and we try to use various techniques to discover what those facts are. Our theory is true if and only if it corresponds to a fact or set of facts. (I'm using the word "fact" to refer to what it is that makes a true proposition true.) The more likely our theories are to match the facts, then the more confident we are in believing our theories. And we may become suspicious of certain theories if we become suspicious either of the nature of the supposed facts our theories reflect or of the nature of our epistemic relation to them. For instance, we do not ordinarily take moral facts, if they existed, to have causal properties. But if we came to think, as some people do, that all existing things have causal properties, then we would become doubtful of moral facts. On the other hand, if we came to think that we cannot learn about anything unless it can have causal effects on our sense perceptions, then we may become doubtful about our capacity to think that we can know anything about moral facts.

I think a moral theory can only be successful if we abandon the usual condition of adequacy. I do not, however, attempt to show that all moral theories must differ from scien-

[20]By saying this, I don't mean to imply that there is agreement among rationalists, sentimentalists, and moral anti-realists about what the conditions of adequacy are. And I am happy to grant that a rationalist, sentimentalist, or anti-realist may have conditions of adequacy different from the ones Rawls attributes to them. Rather, I am drawing a rough caricature in order to make Kant's views clearer.

tific theories in their conditions of adequacy, because some moral theories may share with scientific theories the usual conditions of adequacy. For instance, a devout religious moral theory, perhaps a Christian ethic, might presuppose a divine command theory, which would hold that an ethical proposition is true if and only if it accords with what God commands. In this case, a moral claim has to reflect an objective and independent fact—one which God has actually commanded.

To show that a moral theory has different conditions of adequacy from a scientific theory requires only being able to sketch an acceptable alternative. And I think Rawls and Allison have already pointed the way. This lies in the deep connection Kant makes between morality and reason. Kant is asking if there is anything that reason requires us to do. This question should be no more mysterious, at least metaphysically no more mysterious, than the question of whether there is anything that reason says is in our best interest to do. Kant aims to find something that reason requires us to do. (In the *Groundwork*, he tries to come to this result by examining the very form of acting for a reason, claiming that it follows that there must be something that we must do, and if there is something that we must do, then it must be because the Categorical Imperative exists to give certain commands their legality or normativity. I believe that Kant came to abandon this move.) Thinking of morality in this way means accepting different conditions of adequacy, because we already accept them in other areas.

Let us think of practical reason. It is our human capacity to act for reasons. This involves deliberate reflection and subconscious thought processes. When we look for an adequate theory about how we should behave, we rarely ask for it to satisfy metaphysical demands. One common proposal is that of decision theory. Decision theory holds, roughly speaking, that an action, among a set of given options, is best if and only if its expected value is higher than all the other possible actions (or more accurately, that there is no other action with a higher expected value). All this aside, the point is this: we do not ask of

decision theorists to prove that their thesis reflects certain metaphysical, objective facts. All that it means for an action to be best is that it has a higher expected value than all the other alternatives. We would be misunderstanding what decision theory claims to do if we demand of it that it ought to show that there is a fact about the world which makes that choice best.

Decision theory is a normative theory; it tells to us how we ought to behave. And when we ask if decision theory is true, we are not asking if its axioms reflect an external reality.[21] Instead, we are asking if its axioms are reasonable. We are asking if they are exhaustive. We are asking if its axioms have absurd or contradictory implications. Figuring out the answers to these questions (perhaps there are other questions) is all we need to do in order to find out if decision theory is true. These are our conditions of adequacy for a reasonable theory of practical reason. But one question we do not ask is whether its axioms correspond to an external reality. That question seems to be beside the point.

If we are to look for an alternative to decision theory, we would be looking for axioms that are more reasonable and more exhaustive. I claim that what Kant is doing is very much like trying to find an alternative to decision theory. But he is not trying to find an all-encompassing theory of practical reason. Rather he is trying to find within practical reason whether there is anything that we ought to do, anything that we are required to do, and whether there is anything that we are not allowed to do. What Kant finds is the rule of the Categorical Imperative.

[21]This point is actually much more controversial than I'm making it out to be. Economists, for instance, often use decision theory to predict how people will behave, and insofar as they do so, decision theory is a descriptive project, rather than a normative one. To the extent that it is descriptive, we might demand that its axioms do reflect an external reality—in particular, we might demand that its axioms reflect the actual logic of people's preferences. On that score, there seems to be considerable evidence that people systematically fail to behave according to decision theory's axioms. See Kahneman and Tversky (1979). One way to salvage the merits of decision theory is to think of it as a *normative* theory—one that tells us how we ought to reason, rather than tells us how we do reason. And if it is normative, then it doesn't matter that people don't behave that way. But notice that this move works only because we do not expect a normative theory to reflect an external reality in the same way we expect a descriptive theory to.

We can now re-describe Kant's project in the *Groundwork*. He was convinced, for the reasons I set out in the first chapter, that morality, if it exists, must essentially be based on reason. So in that work, Kant tries to do two things. The first is establish that the very concept of obligation entails the concept of the Categorical Imperative. We saw how Korsgaard tried to give an account of that argument above, and I expressed, in an earlier chapter, my doubts about the success of such an attempt. The second, as Korsgaard puts it, is to show that obligations really exist (1996, 67). In the *Groundwork*, Kant tries to do this by showing that our being rational entails our being autonomously motivated, and that somehow means we are capable of being moved by obligations. I expressed even more doubts about such an approach.

In the second *Critique* Kant abandons the second approach but still accepts the first half of the argument. In 5: 21–27, Kant tries to show again that only purely formal characteristics of a maxim (or motive) can ground an obligation, and any empirically based maxim cannot do so, because it will always be ultimately optional. He then tries to show that the relevant formal characteristic must be the mere form of a law, which can be characterized as the Categorical Imperative. I had argued that much of this argument was compelling, save for the final step, the move from the mere form of law to the Categorical Imperative.

But he abandons the second approach in favour of the approach of the fact of reason. So instead of trying to show that obligations are real because we can be motivated by our own will, we can instead start with the fact of reason, the claim that what makes an action right can, and sometimes, does guide our actions, and that shows that we are autonomous. There are two questions about this: How does this justify morality? How does it show us our autonomy?

Let us answer the first question first. For Kant, what makes an action right or wrong (or more accurately, permissible or impermissible) is whether it passes the CI-procedure. And given my account of what it means to act from the motive of duty, the fact of reason is

best understood as the following claim:

> The justificatory standard of the CI-procedure can have a decisive and appropriate effect on whether a human being performs an action. (By "appropriate", I mean that if it fails, S refrains from performing that action.)

The fundamental moral principle, for Kant, is essentially a justificatory standard. And, as I argued above, morality is essentially about justification. Since we do not, in general, require that our justificatory standards reflect an external reality, we would be remiss to try to show that it does. And it would be nonsensical of a sceptic to require that morality do so.

That may deflect the questions of certain kinds of moral anti-realists, but we still have to do two things. One, we have to show that the CI-procedure is the right justificatory standard for action. There are, after all, alternatives in the form of alternative normative moral theories and various forms of decision theory. Kant tries to show that the CI-procedure is the right standard by deriving it from the concept of obligation, because the concept of obligation is the concept inherent in the idea of justification for action. So if one accepts his analysis of obligation in chapter 1 of the *Groundwork* and in 5:21–27 of the second *Critique*, then one accepts that the CI-procedure is the ultimate justificatory standard. One may, however, find this argument unconvincing, and as I argued in my second chapter, it is difficult to make the argument work.

There is an alternative way of endowing the CI-procedure with its status as a justificatory standard. But in exploring it, we leave the project of interpreting Kant behind. What we would do is argue that we find the principle exhibited in our moral life, and that it is present in the stable result of a process of reflective equilibrium. The process of reflective equilibrium consists of working between our particular moral judgments and the various moral principles we take to guide them. Being imperfect human beings, we will undoubtedly have inconsistencies and incoherencies among our principles and considered

judgments. We revise our judgments and principles, in an ongoing project to achieve a certain coherence and stability, a state of equilibrium. And once we reach that equilibrium, we are justified in accepting the resultant new considered judgments and principles. Thus, justifying the CI-procedure involves finding it at the core of our principles in the resultant equilibrium. Rawls seems to endorse this approach. He thinks the CI-procedure mirrors what we value most: our equality, freedom, reasonableness, and personhood. I'm uncertain of the extent to which he thinks this is Kant's view. And I also have a serious misgiving about this approach. It doesn't establish the CI-procedure as the ultimate justificatory standard until we actually go through and complete the process of reflective equilibrium. If we have completed such a process, I would have to show the process and show that the CI-procedure is the ultimate principle in our moral thought in order to take it as the justificatory standard for action. But if we haven't yet reached such an equilibrium, we cannot predict beforehand what our principles will be. In sum, reflective equilibrium is a nice goal, but we do not know what it will justify until we reach it.

Two, we have to show that the CI-procedure does have this effect on human beings; that is, that the fact of reason is indeed a fact. How does Kant show this? In short, he doesn't. He asserts it. After all, such a claim is a contingent fact, and it is akin to an empirical claim. It is contingent whether people do or do not, whether people ever have or have not, been appropriately affected by the CI-procedure. How could we determine if such a principle does have such an effect? We can only do so much in examining a person's thought processes. We can look at a person's action, and we may try to infer from his action and his circumstance that there could be no reason for his doing what he did unless he was motivated by the CI-procedure. But even in so doing, we are at best guessing. The motivations of human beings are so complex and hidden from us that we do not know what it is that genuinely moves a person to act the way he did. We could, instead, introspect, and see that we test to see if our maxims are universalizable and thus act from them. We

could even make a conscious effort to *choose* to act in accordance with the CI-procedure. It would seem that by deliberately, consciously making such a choice, we are indeed clearly acting from the motive of duty. (Of course, it doesn't have to be so deliberate for us to have acted from the motive of duty.) But things are not so clear. All this provides evidence for us to think that the fact of reason is true, but none of this guarantees that we know it.

After all, introspection for Kant is a shady process, and we are never certain about our motives. This is why Kant says that "no exact observance of it can be found in moral experience." (5:47) But if we cannot *look and see* if we are acting from the motive of duty, how do we know that we are? We cannot, alternatively, derive this fact of reason from reason itself, because it is a contingent claim and it is a fact about reason's influence on us human beings. Kant's unhelpful answer to the question of how we know is that we are "a priori conscious" of it (5: 47). It seems that we can neither prove it empirically nor derive it analytically, yet, according to Kant, we nonetheless know it to be a fact. But if we don't know how we know it, can we say that we know it?

That depends. We might say that knowledge requires knowing how we know something, and thus say that we don't know the fact of reason. Or we could reject that requirement, and say that maybe we do know it but we don't know how we know it. Kant makes the second choice; he says that we do know it but also that it is impossible for us to know how the fact of reason could be true.

In a previous chapter, I pointed out that Allison said that knowing how the fact of reason could be true is tantamount to knowing how practical reason is possible, which is to come to the bottom of a faculty and that means, for Kant, we can not know how it is possible. This isn't exactly right, because practical reason only means we can act for reasons. It doesn't show that we have obligations too. But I argued that the capacity to act for reasons does not show us that we have moral obligations when I argued that Kant's argument in *Groundwork III* fails at the step from negative freedom to positive freedom. It is precisely

on this point where I think Kant changed his mind. To prove that we have obligations is simply to show to us that we do, and that is what the fact of reason attests by claiming that we accept the Categorical Imperative as guiding.

The second question: how does the fact of reason show us our autonomy? As I argued in chapter 2, an agent is autonomous if and only if (1) an agent can act in accordance with a principle that does not depend on any of her desires, and (2) the principle in question comes from the agent herself. The fact of reason implies that (1) is satisfied, because it asserts that we can act in accordance with the Categorical Imperative, which does not depend on any of our desires. The question now is whether (2) is satisfied—whether the Categorical Imperative comes from the agents themselves. This is a difficult question to answer, but if the Categorical Imperative is a requirement of reason, and if it is synthetic a priori, as Kant thinks it is, then there are good Kantian reasons to think that it must somehow come from the agents themselves. How this might work exactly, I do not know.

This leads us to another question: how does the fact of reason show to us our *transcendental* freedom? What I've said above seems to show to us our freedom, but not necessarily our transcendental freedom. But how we are to understand Kant's insistence that through being aware of the moral law, we can cognize our own transcendental freedom? Kant writes,

> Therefore, that unconditioned causality and the capacity for it, freedom, and with it a being (I myself) that belongs to the sensible world but at the same time to the intelligible world, is not merely *thought* indeterminately and problematically (speculative reason could already find this feasible) but is even *determined with respect to the law of its causality* and *cognized* assertorically; and thus the reality of the intelligible world is given to us, and indeed as *determined* from a practical perspective, and this determination, which for theoretical purposes would be *transcendent* (extravagant), is for practical purposes *immanent*. (CPrR 5: 105)

But one of the lessons of the first *Critique* was that we could not know anything about the noumenal or transcendental realm.[22] So how is it that we can cognize our transcendental freedom? We can think of Kant's answer in two parts. First, it is Kant's contention that no freedom other than transcendental freedom will do for moral responsibility. Second, moral responsibility requires some kind of freedom. So once we recognize the authority of the moral law, we must also accept our transcendental freedom.

Kant places rather stringent requirements on moral responsibility. In places, he contends that our moral responsibility requires freedom from the causal order of the universe. He writes,

> If I say of a human being who commits a theft that this deed is, in accordance with the natural law of causality, a necessary result of determining grounds in preceding time, then it was impossible that it could have been left undone; how, then, can appraisal in accordance with the moral law make any change in it and suppose that it could have been omitted because the law says that it ought to have been omitted? That is, how can that man be called quite free at the same point of time and in regard to the same action in which and in regard to which he is nevertheless subject to an unavoidable natural necessity? (CPrR 5: 95–6)

Kant goes on to criticize classical compatibilists for trying to locate freedom *within* a deterministic world and for trying to revise our conception of freedom such that it doesn't require a capacity to have done otherwise. Moral responsibility, for Kant, thus seems to require a capacity to have done otherwise and for us to have been the original source of an action. If the action has its determining causes in the past prior to our action, then we could hardly be said to have caused the action. But Kant believes that we cannot find evidence of anything that would meet these two requirements for freedom in the empirical world, partly

[22]Of course, it may be contended that Kant thinks we know *something* about noumena; for instance, it might be contended that Kant believes that there are noumena. Moreover, some contend that his claims regarding the deduction of categories are claims about noumena, which seem to be inconsistent with his claims about the limits of what we can know about noumena. See Strawson (1990). But most seem to agree that Kant's transcendental idealism rules out knowing anything about noumena's causal power. See Pereboom (1991). And that is the relevant point here. Being transcendentally free means that we can cause actions, and that we can be at the beginning of a causal chain independently of a deterministic chain of causes in the empirical world. But Kant insists on our transcendental freedom anyway.

because he thinks determinism is utterly comprehensive in its reach. And partly because we can have no intuition of our freedom.

This last point is worth dwelling upon. Kant thinks that the only intuitions we have are sensuously conditioned.[23] And as such, we cannot have any intuition that these two requirements are met by anything. But what this implies is that more common appeals to examples in our lives as evidence of our freedom fails. For instance, one might think that because I can apparently choose between soup or salad for dinner, and there is no apparent deciding factor of one over the other, that I am free. But for Kant this will not do. Just because it seems to me that I can make a choice between this or that does not indicate, for Kant, an independence from the empirical conditions of the world. This may be a mere seeming, which need not correlate to anything substantive. Thus, only the moral law, understood as the Categorical Imperative, will do. This is because the moral law makes demands on us that do not depend on empirical conditions. But what makes this difficult is that we don't have an intuition of the moral law either. We cannot have any direct access to, or sensuous intuition of, the moral law. All we have instead is the fact of reason, the fact that we accept, at least implicitly, the Categorical Imperative as a decisive influence on our actions. Kant, I suspect, must also acknowledge the possibility that the fact of reason might also be an illusion. But if we recognize its authority, we must also recognize and accept what it requires; in particular, it requires our transcendental freedom.

There is a looming difficulty I do not wish to resolve here, but I do need to make it clear and less threatening. And that problem is to make his claims here consistent with Kant's belief that determinism is utterly comprehensive in its reach, including all our human actions, and that as such it rules out the possibility of our freedom. Kant writes,

> Now, since time past is no longer within my control, every action that I perform

[23] Kant writes, "The capacity (receptivity) for receiving representations through the mode in which we are affected by objects, is entitled sensibility. Objects are given to us by means of sensibility, and it alone yields us intuitions" (A 19/B 33).

> must be necessary by determining grounds *that are not within my control*, that is, I am never free at the point of time in which I act.... For, at every point of time I still stand under the necessity of being determined to action by *that which is not within my control*, and the series of events infinite a parte priori which I can only continue in accordance with a predetermined order would never begin of itself: it would be a continuous natural chain, and therefore my causality would never be freedom (CPrR 5:94–5).

What this seems to imply is that Kant endorses incompatibilism—the claim that freedom and determinism are incompatible. But at the same time, Kant endorses both freedom and determinism. Resolving this contradiction depends, in part, on his account of transcendental idealism, and his view that determinism applies only to the causal empirical world and not to things in themselves. Because we are also noumenal beings, we can think ourselves to be free in the noumenal world. But this reply only goes so far, because as you can see from one of the quotations above, Kant says that the "unconditioned causality and the capacity for it, freedom, and with it a being (I myself) ... belongs to the sensible world" and the intelligible world (CPrR 5: 105). Thus freedom, for Kant, must exist within the sensible world too, and not just in the intelligible world, even though determinism is true in the sensible world and that it precludes freedom.

I do not need to resolve this problem here. If Kant's theory could resolve this problem, he has been able to satisfy all that one might ever want in the free will debate: we can have determinism and all its implications, and we can have all that we've ever wanted from free will. My argument regarding the fact of reason is largely independent of this problem. But I needed to explain why morality requires transcendental freedom, and why only the moral law, understood as the Categorical Imperative, can reveal it to us. If we accept the moral law as authoritative, as the fact of reason claims, then we must also accept our transcendental freedom. Figuring out how to do that while maintaining Kant's incompatibilism is a separate matter.

4.5 Conclusion

This project has been an attempt, in part, to explain Kant's move from his argument for the justification of morality in the *Groundwork* to his argument in the second *Critique*. So to sum up, in the *Groundwork*, Kant argued from our theoretical rationality to our having a will and to our negative freedom, from our negative freedom to our positive freedom, and from our positive freedom to our standing under the moral law. In broader outlines, he argued from our capacity for theoretical reason to our capacity for practical reason, and from practical reason to the existence of obligations, and from the concept of obligation, the moral law. In the second *Critique*, instead of arguing for our practical rationality, it is his starting place. That work, after all, is an investigation of our faculty of practical reason. He argued, as he did in the *Groundwork*, that the concept of obligation (or the capacity of reason to have *decisive* influence on our behaviour) gives us the Categorical Imperative. But what he rejects in the second *Critique* is, I suspect, being able to show our capacity for practical reason implies that we have obligations; that is, he no longer thought that being able to act for reasons implies that we take some reasons as having decisive influence on our behaviour. He replaces this step with the fact of reason: the claim that we do take a certain kind of reason, the CI-procedure, to be decisive on our behaviour. Since we do accept the CI-procedure as authoritative, we have obligations, and as such, morality is real.

My approach in this final chapter has been this: to make as much of what Kant says as true as possible. One may find that what I've said goes far too against the grain of what Kant says in some places, particularly on what he means by acting from the motive of duty. But I believe that I have been able to preserve the important and worthwhile aspects of his claims and to show how they can be preserved in light of certain problems. I argued that some incoherencies in Kant's account, given by Korsgaard's interpretation, could be relieved by re-thinking what it means to act from the motive of duty.

I found it necessary to give, what I take to be, a novel account of what the fact of reason is, emphasizing its close connection to Kant's idea of acting from the motive of duty, and how I took it to justify morality. Once we assume that we can act for reasons, all we need to do to show that morality is real is to show that there is a principle that we, in fact, accept as ultimately justificatory. This principle for Kant is the CI-procedure and I tried to show how we ought to think of it as a justificatory standard. Thus, once we accept that we can act for reasons and that we do regard the CI-procedure as a guiding principle on our behaviour, we have shown that morality exists. But here are two questions that I have been unable to answer: Do we actually find the CI-procedure in our thought? And supposing we do find it in our thought, does that mean that it is justified? Kant believes that we do find it in our thought, but that is contentious. I find it easier to believe that our ordinary thought reveals a related, but weaker principle—something akin to "if I take a reason to justify my action in particular circumstances, I must understand it to justify anybody's action in relevantly similar circumstances", or perhaps equivalently, "In deciding how to act, I must not make an exception of myself". But the more substantive Categorical Imperative, as embodied in the CI-procedure, is harder to find in our ordinary moral thought. As for the second question, I believe Kant answers it by showing that the CI-procedure follows from the concept of accepting an obligation (or the concept of a reason that has decisive influence). But I also expressed scepticism of this argument. Nonetheless, I think I have been able to give a plausible account of the fact of reason, which shows its hitherto unseen, intimate connection to the concept of acting from the motive of duty, its role in justifying morality, and to some extent, how it reveals to us our transcendental freedom.

Bibliography

Adams, Robert M. *Leibniz. Determinist, Idealist, Theist*. Oxford University Press, 1994.

Allison, Henry E. "Morality and Freedom: Kant's Reciprocity Thesis." *The Philosophical Review* 95, 3: (1986) 393–425.

———. *Kant's Theory of Freedom*. Cambridge University Press, 1990.

Ameriks, Karl. *Interpreting Kant's Critiques*. Oxford University Press, 2003.

Arpaly, Nomy. "Moral Worth." *The Journal of Philosophy* 99, 5: (2002) 223–245.

———. *Merit, Meaning, and Human Bondage: An Essay on Free Will*. Princeton University Press, 2006.

Aune, Bruce. *Kant's Theory of Morals*. Princeton University Press, 1979.

Baron, Marcia. "The Alleged Moral Repugnance of Acting from Duty." *The Journal of Philosophy* 81: (1984) 197–220.

Beck, Lewis White. *A Commentary on Kant's Critique of Practical Reason*. University of Chicago Press, 1963.

Bennett, Jonathan. "The Conscience of Huckleberry Finn." *Philosophy* 49, 188: (1974) 123–134.

Bermúdez, José Luis. *Decision Theory and Rationality*. Oxford University Press, 2009.

Broome, John. "Normative Requirements." *Ratio* 12: (1999) 398–419.

Brown, Charlotte. "From Spectator to Agent: Hume's Theory of Obligation." *Hume Studies* 20, 1: (1994) 19–36.

Carroll, Lewis. "What the Tortoise Said to Achules." *Mind* 4, 14: (1895) 278–280.

Chisholm, Roderick M. *The Problem of the Criterion*. Marquette University Press, 1973.

Copp, David. "Belief, Reason, and Motivation: Michael Smith's 'The Moral Problem'." *Ethics* 108, 1: (1997) 33–54.

Dancy, Jonathan. *Ethics Without Principles*. Oxford University Press, 2004.

Daniels, Norman. "Reflective Equilibrium." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2008. Fall 2008 edition.

Darwall, Stephen. *Philosophical Ethics*. Westview Press, 1998.

Dennett, Daniel. *Freedom Evolves*. Penguin Books, 2003.

Descartes, René. *Meditations on First Philosophy: With Selections from the Objections and Replies*. Cambridge University Press, 1996. Trans. John Cottingham.

Dietrichson, Paul. "What Does Kant Mean By 'Acting From Duty'?" In *Kant: A Collection of Critical Essays*, edited by Robert Paul Wolff, Anchor Books, 1967, 314–330.

Dworkin, Ronald. "Objectivity and Truth: You'd Better Believe It." *Philosophy & Public Affairs* 25, 2: (1996) 87–139.

Frankfurt, Harry G. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66, 23: (1969) 829–839.

Gardner, Sebastian. *Routledge Philosophy Guidebook to Kant and the Critique of Pure Reason*. Routledge, 1999.

Goldman, Alvin I. "A Causal Theory of Knowing." *The Journal of Philosophy* 64, 12: (1967) 357–372.

Harman, Gilbert. *The Nature of Morality: An Introduction to Ethics*. Oxford University Press, 1977.

Henrich, Dieter. *The Unity of Reason: Essays on Kant's Philosophy*. Harvard University Press, 1994. Edited by Richard L. Velkley.

Herman, Barbara. *The Practice of Moral Judgment*. Harvard University Press, 1993.

Hill, Thomas E. *Dignity and Practical Reason in Kant's Moral Theory*. Cornell University Press, 1992.

Hintikka, J. "'Knowing That One Knows' Reviewed." *Synthese* 21, 2: (1970) 141–162.

Hume, David. "An Enquiry Concerning the Principles of Morals." In *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, edited by LA Selby-Bigge & PH Nidditch, Clarendon Press, 1975. 3rd edition.

———. *A Treatise of Human Nature*. Oxford University Press, 1978, 2nd edition. Edited by L. A. Selby-Bigge & P. H. Nidditch.

Jolley, Nicholas. *Leibniz*. Routledge, 2005.

Kagan, Shelly. *The Limits of Morality*. Oxford University Press, 1989.

Kahneman, Daniel, and Amos Tversky. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47, 2: (1979) 263–291.

Kane, Robert. *The Significance of Free Will*. Oxford University Press, 1996.

———. *A Contemporary Introduction to Free Will*. Oxford University Press, 2005.

Kant, Immanuel. *Critique of Pure Reason*. St. Martin's Press, 1965. Trans. Norman Kemp Smith.

———. "What is orientation in thinking." In *Kant: Political Writings*, edited by Hans Reiss, Cambridge University Press, 1991. Trans. H.B. Nisbett.

———. "On a supposed right to lie because of philanthropic concerns." In *Grounding for the metaphysics of morals*, edited by James W. Ellington, Hackett Publishing, 1993.

———. *Critique of Practical Reason*. Cambridge University Press, 1997. Trans. Mary Gregor.

———. *Groundwork of the Metaphysics of Morals*. Cambridge: Cambridge University Press, 1998a. Trans. Mary Gregor.

———. *Religion within the Boundaries of Mere Reason*. Cambridge University Press, 1998b. Trans. Allen Wood & George di Giovanni.

Kerstein, Samuel. "Deriving the Supreme Principle of Morality from Common Moral Ideas." In *The Blackwell Guide to Kant's Ethics*, edited by Thomas E. Hill, Blackwell Publishing, 2005, 121–137.

———. "Reason, Sentiment, and Categorical Imperatives." In *Contemporary Debates in Moral Theory*, edited by James Dreier, Blackwell Publishing, 2006, 129–143.

Korsgaard, Christine. *Creating the Kingdom of Ends*. Cambridge University Press, 1996.

———. "The Normativity of Instrumental Reason." In *Ethics and Practical Reason*, edited by Garrett Cullity & Berys Gaut, Clarendon Press, 1997.

———. "Introduction." In *Groundwork of the Metaphysics of Morals*, Cambridge University Press, 1998. Trans. Mary Gregor.

Leibniz, Gottfried. *The Monadology and Other Essays*. Oxford University Press, 1965. Trans. Robert Latta.

———. *Theodicy: Essays on the Goodness of God, the Freedom of Man, and the Origin of Evil*. Open Court Publishing, 1985. Trans. E. M. Huggard.

———. "Discourse in Metaphysics." In *Philosophical Essays*, edited by Roger Ariew & Daniel Garber, Hackett Publishing, 1989a, 35–68.

———. "On Freedom." In *Philosophical Essays*, edited by Roger Ariew & Daniel Garber, Hackett Publishing, 1989b, 94–97.

———. *Philosophical Essays*. Hackett Publishing, 1989c. Trans. Roger Ariew & Daniel Garber.

———. *New Essays on Human Understanding*. Cambridge University Press, 1996. Trans. Peter Remnant & Jonathan Bennett.

Leibniz, Gottfried, and Antoine Arnauld. *The Leibniz-Arnauld Correspondence*. Manchester University Press, 1967. Trans. H.T. Mason.

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

Miller, Alexander. *An Introduction to Contemporary Metaethics.* Polity Press, 2003.

———. *Philosophy of Language*. Routledge, 2007, 2nd edition.

Moore, George Edward. "A Reply to My Critics." In *The Philosophy of G. E. Moore*, edited by P.A. Schilpp, Northwestern University Press, 1942, 535–677.

———. "Proof of an External World." In *Epistemology: An Anthology*, edited by Ernest Sosa & Jaegwon Kim, Blackwell Publishing, 2000, 24–26.

Norton, David Fate. "Hume, Human Nature, and the Foundations of Morality." In *The Cambridge Companion to Hume*, edited by David Fate Norton, Cambridge University Press, 1993, 148–181.

O'Neill, Onora. *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge University Press, 1989.

———. *Bounds of Justice*. Cambridge University Press, 2000.

Paton, H. J. *The Categorical Imperative*. Harper Torchbooks, 1967.

Pereboom, Derk. "Is Kant's Transcendental Philosophy Inconsistent?" *History of Philosophy Quarterly* 8, 4: (1991) 357–372.

———. "Hard Incompatibilism." In *Four Views on Free Will*, edited by John Martin Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas, Blackwell Publishing, 2007a, 85–125.

———. "Kant on Transcendental Freedom." *Philosophy and Phenomenological Research* 73, 3: (2007b) 537–567.

Plato. *Five Dialogues: Euthyphro, Apology, Crito, Meno, Phaedo*. Hackett Publishing, 2002. Trans. G. M. A. Grube & John M. Cooper.

Prichard, H. A. "Does Moral Philosophy Rest On a Mistake?" *Mind* 21, 81: (1912) 21–37.

Rawls, John. "Kantian Constructivism in Moral Theory." *The Journal of Philosophy* 77, 9: (1980) 515–572.

———. "Themes in Kant's Moral Philosophy." In *Kant's Transcendental Deductions*, Stanford University Press, 1989, 89–113.

———. *Lectures on the History of Moral Philosophy*. Harvard University Press, 2000. Edited by Barbara Herman.

Rotenstreich, N. "Will and Reason: A Critical Analysis of Kant's Concepts." *Philosophy and Phenomenological Research* 46, 1: (1985) 37–58.

Rousseau, Jean-Jacques. *The Social Contract and The First and Second Discourses*. Yale University Press, 2002. Edited by Susan Dunn.

Schneewind, J. B. *The Invention of Autonomy*. Cambridge University Press, 1998.

Schroeder, Mark. "The Hypothetical Imperative?" *Australasian Journal of Philosophy* 83, 3: (2005) 357–372.

Schwartz, Jeremy. "Do Hypothetical Imperatives Require Categorical Imperatives?" *European Journal of Philosophy* 18, 1: (2008) 84–107.

Shafer-Landau, Russ. *Moral Realism: A Defence*. Oxford University Press, 2003.

———. "Ethics as Philosophy: A Defense of Ethical Nonnaturalism." In *Metaethics after Moore*, edited by Terry Horgan & Mark Timmons, Oxford University Press, 2006, 209–232.

Singer, Peter. "Famine, Affluence, and Morality." *Philosophy & Public Affairs* 1, 3: (1972) 229–243.

Smith, Michael. *The Moral Problem*. Blackwell Publishing, 1994.

Stocker, Michael. "The Schizophrenia of Modern Ethical Theories." *The Journal of philosophy* 73, 14: (1976) 453–466.

Stratton-Lake, P. *Kant, Duty and Moral Worth*. Routledge, 2004.

Strawson, P. F. *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*. Routledge, 1990.

Street, Sharon. "Constructivism about reasons." In *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, Oxford University Press, 2008, volume 3, 207–45.

———. "What is Constructivism in Ethics and Metaethics?" *Philosophy Compass* 5, 5: (2010) 363–384.

Sullivan, Roger J. *Immanuel Kant's Moral Theory*. Cambridge University Press, 1989.

———. *An Introduction to Kant's Ethics*. Cambridge University Press, 1994.

Sussman, David G. *The Idea of Humanity: Anthropology and Anthroponomy in Kant's Ethics*. Routledge, 2001.

Svavarsdóttir, Sigrún. "How Do Moral Judgments Motivate?" In *Contemporary Debates in Moral Theory*, edited by James Dreier, Blackwell Publishing, 2006.

Twain, Mark. *Adventures of Huckleberry Finn*. Signet Classic, 1997.

Unger, Peter K. *Living High and Letting Die: Our Illusion of Innocence*. Oxford University Press, 1996.

Williams, Bernard. *Morality*. Cambridge University Press, 1972.

———. *Ethics and the Limits of Philosophy*. Harvard University Press, 1985.

Williamson, Timothy. *Knowledge and Its Limits*. Oxford University Press, 2000.

Wolf, Susan. "Moral Saints." *The Journal of Philosophy* 79, 8: (1982) 419–439.

Wood, Allen. *Kant*. Blackwell Publishing, 2005.

CURRICULUM VITAE

Name	Kenneth K. H. Chung
Post-secondary Education and Degrees	McGill University Montréal, Québec, Canada 1997–2000 B.A.
	The University of Saskatchewan Saskatoon, Saskatchewan, Canada 2000–2002 M.A.
	The University of Western Ontario London, Ontario, Canada 2002–2010 Ph.D.
Honours and Awards	Ontario Graduate Scholarship 2004–2005, 2006–2007
	Social Sciences and Humanities Research Council (SSHRC) Doctoral Fellowship 2005–2006
Related Work Experience	Lecturer The University of Western Ontario 2004–2010
	Teaching Assistant The University of Western Ontario 2002–2004