MORAL RESPONSIBILITY AND THE NATURAL ORDER

by

Katy Allen

A thesis submitted to the Department of Philosophy
In conformity with the requirements for
the degree of Master of Arts

Queen’s University
Kingston, Ontario, Canada
(September, 2007)

Copyright © Katy Allen, 2007
Abstract

This thesis examines Kantian conceptions of freedom. Beginning with Kant himself, I show how Kant’s response to Hume concerning the rational justification of causal judgments results in his claim that the sensible world is governed a priori by causal principles. Kant’s moral philosophy, however, requires a robust conception of freedom for moral agency to be possible. These two features leave Kant in an apparent contradiction, for it is unclear how we, as members of the physical, causal world, can be truly free if all events are governed by causal laws. I show that Kant’s solution to this contradiction lies in an important aspect of his transcendental idealism: the noumenal/phenomenal distinction. I argue, further, that his solution is problematic due to the fundamentally unknowable quality of the noumenal realm, wherein freedom is located. John McDowell’s Mind and World is introduced as an alternative to the extreme Kantian dualism between noumena and phenomena, while remaining within a broadly Kantian framework. Like Kant, McDowell locates our freedom in our ability to operate through reason, though unlike his predecessor, he situates “the space of reasons” within nature. This becomes possible by extending our conception of nature to include a “second nature”, thus making our initiation into the space of reasons—into the realm of freedom—a natural process. Remaining Kantian in spirit, however, McDowell’s account inherits a problematic Kantian feature. He maintains the distinction between two modes of intelligibility—between naturalistic and rational modes of explanation—thus leaving room for a hard-nosed naturalist to question the autonomy of the latter. I argue that Peter Strawson’s proposal in “Freedom and Resentment” is able to assuage this worry in
McDowell’s otherwise plausible model. In it, Strawson provides an account of why the autonomy of rational explanations can never be undermined by purely naturalistic explanations, even in the face of a theoretical conviction in determinism. Strawson argues that our “personal reactive attitudes” (like gratitude and resentment)—attitudes that express our commitment to a moral life and are representative of our functioning within the space of reasons—could never be undermined by the truth of determinism, and this reveals the extent to which our conception of ourselves as rational agents is immune from assault by the determinist. The result is a compelling form of compatibilism that persuasively retains the space of reasons without appeal to Kantian noumenalism.
Acknowledgements

Thank you endlessly to my extraordinarily helpful and encouraging supervisor, Professor David Bakhurst.

Thank you to my fellow Kantian and personal chauffeur, Anthony Bruno.

Thank you to my four ladies (and one man): Jessie, Kiri, Ali, Siobhan, & Erik.

Most of all thank you to the two coolest human beings I have yet to meet, and my personal heroes: Mom and Dad.
# Table of Contents

Abstract..................................................................................................................................................... ii

Acknowledgements ................................................................................................................................... iv

Table of Contents ....................................................................................................................................... v

Chapter 1.................................................................................................................................................. 1

Chapter 2.................................................................................................................................................. 27

Chapter 3................................................................................................................................................ 46

Bibliography ............................................................................................................................................. 67
(i) Introduction

This thesis grapples with one of the most enduring problems of modern philosophy: the possibility of freedom in the natural world. It is widely believed that moral responsibility is possible only if human beings are genuinely free agents in control of their actions, for we can be held accountable for our acts only in cases where we could have acted otherwise. Yet how is it possible to reconcile this view of freedom with the fact, now rarely disputed, that human beings are creatures of the natural world, the events of which are wholly dictated by causal law? One figure for whom these issues are especially pressing is Kant, whose work provides an ingenious attempt to portray human beings as free rational agents notwithstanding their status as inhabitants of the empirical world subject to its causal laws. I begin this thesis by considering the philosophical framework in which Kant develops his conception of freedom. I argue that Kant’s view is ultimately unsatisfactory, but that a more compelling account of the place of freedom in nature can be developed by exploring the work of two contemporary philosophers who are much inspired by him: John McDowell and Sir Peter Strawson.

(ii) Background: Hume on causation

This chapter explores Kant’s attempt to reconcile the causal, physical world with freedom—the existence of which, Kant believes, is a necessary condition of moral action and responsibility. As Kant’s conception of causation is customarily seen as a response to Hume’s considerations on the same topic, section (ii) sketches Hume’s famous
discussion of causation. Section (iii) then explores Kant’s response to Hume, examining in particular the argument put forth in the “Second Analogy” of the *Critique of Pure Reason*. Sections (iv) and (v) then address the place of freedom in Kant’s conception of morality, before I turn my attention to the problem of reconciling his ethics with his metaphysics. The question is this: How can we act in accordance with the moral law (and so be seen as morally capable and responsible agents) if our existence in the physical world is causally determined? In answering this question, I appeal to Kant’s arguments for freedom presented in Chapter III of the *Groundwork of the Metaphysic of Morals* and the *Critique of Practical Reason*. In section (vi) I argue that Kant’s attempt to reconcile causation and freedom—through his appeal to the phenomenal/noumenal distinction—remains deeply flawed.

Hume’s two main arguments on causation, which are meant to show that causal judgments are not grounded in reason, are presented in Sections IV and VII of his *Enquiry Concerning Human Understanding*, which represents a somewhat simplified version of the arguments originally developed in Book One of his *Treatise of Human Nature*. In Section IV of the *Enquiry*, Hume states that the conditions for rationally justifying a causal judgment must appeal to one of the two types of human reasoning: “relations of ideas” or “matters of fact” (Hume, 1977: 15). The former are judgments that can be justified *a priori*, such as mathematical or logical propositions, merely by considering the meaning or content of the ideas that comprise them; the latter, matters of fact, are justified empirically (so *a posteriori*). For Hume, a causal judgment—e.g. that *this* event will necessarily lead to *that* event—cannot be justified *a priori*, since their
truth does not follow merely from the meaning of the ideas it contains. Adam, Hume
tells us, could not have known that water might drown him merely by reflecting on the
idea of water (Hume, 1977: 17). Further, it is the mark of judgments true in virtue of
relations of ideas that their negation either is or entails a contradiction, but this is never
true of causal judgments.

Having shown that causal judgments cannot be proven *a priori*, Hume concludes
that our knowledge of cause and effect must be based in experience—matters of fact. My
belief then, that A will cause B, must be grounded in the fact that I have prior experience
of events of type A causing events of type B. This example reflects reasoning based on
induction, suggesting that we can base future causal judgments on what we have always
observed in the past. For induction to be a valid form of reasoning, however, we need to
appeal to some sort of *principle of the uniformity of nature* that establishes that the future
will indeed resemble the past. Hume’s insight into this is to assert that no such principle
can be rationally justified.

Firstly, such a principle cannot not be justified *a priori* since its denial is not
contradictory. It is logically conceivable to suppose that the world is not governed by
such a principle. Second, if we try to justify the principle by appeal to experience, we fall
into a vicious circle. We cannot argue that the future will continue to resemble the past
because, in the past, the future has always resembled the past. This is to assume what we
are trying to prove. Hume’s conclusion then, in Section IV, is that causal judgments
cannot be based in reason—either *a priori* or *a posteriori*. Though this argument has
merely demonstrated Hume’s negative answer regarding causation (so it has told us what
causation is not), let me now turn my attention to his second argument presented in Section VII, in order to clarify his position and present his positive account on causation.

Hume begins Section VII of the *Enquiry* by defining mental content as that which is made up of “impressions” and “ideas” (Hume, 1977: 41). He defines the former, impressions, as the immediate objects of sensory input. The latter, ideas, are characterized as indirect copies of earlier experienced impressions, thus either remembered or synthesized sense perceptions. Though ideas may combine with other ideas to form more complicated ones, they can always be broken down into earlier-experienced impressions. Every idea, therefore, has as its foundation one or multiple sense impressions from which it originated.

Thus to make sense of the complex idea of causation, it must be possible to trace it back to certain earlier formative impressions. Hume argues that the components that form our idea of causation are threefold: spatio-temporal contiguity, priority, and a necessary connection (Hume, 1977: 14). Thus to say that event A causes event B, we must not only observe that events A and B exist in a spatially contiguous fashion with the former always preceding the latter, but also that there exists some sort of necessity that compels the occurrence of the second event. This idea of a necessary connection is critical, since there are certainly events that fulfill the first two criteria, though we would not say that they were causally related. For example, illnesses may present themselves in this way: a certain bacterial infection may begin in a patient as a fever, followed by a cough, followed by a rash. This case fulfills the criteria of spatio-temporal contiguity and priority, though it would be incorrect to say that the fever caused the cough, and the
cough caused the rash. Instead, it is that the presence of bacteria caused all the symptoms. Thus the need to introduce the criterion of necessary connection.

Hume contends that although we may properly observe spatio-temporal contiguity and priority, we never genuinely observe necessary connections between events. If he can show this, this intensifies his reasons for thinking that causal judgments are not empirically justified. Referring to Hume’s famous billiard ball example as proof for this latter claim, Hume asserts that although we see event A (one billiard ball in motion hitting another billiard ball) and event B (the hit billiard ball moving upon impact), we do not observe a genuine necessary connection between these two events, but only priority and spatio-temporal contiguity. As Hume puts it in the *Enquiry*:

> When I see, for instance, a billiard ball moving in a straight line towards another; even suppose motion in the second ball should by accident be suggested to me, as the result of their contact or impulse, may I not conceive that a hundred different events might as well follow from the cause? May not both these balls remain at absolute rest? May not the first ball return in a straight line, or leap off from the second in any line or direction? All these suppositions are consistent and conceivable. Why then should we give the preference to one, which is no more consistent or conceivable than the rest? (Hume, 1977: 18)

So if we cannot, as Hume asserts, locate the impression of a necessary connection between observable events, where does this idea come from? Hume’s proposal is that our idea of a necessary connection originates firstly from experiencing events in terms of priority and spatio-temporal contiguity. Over time, he claims, we become accustomed to experiencing certain events in such a way—what Hume refers to as a ‘constant conjunction’ between particular events (Hume, 1977: 46). After multiple instances, Hume believes that our minds inevitably come to form a connection between the two
events. This connection, Hume says, is interpreted by humans as a necessary connection existing in the external world. Thus we come to posit causation as a genuine feature of the sensible world.

Hume, however, takes this last conclusion to be mistaken. His claim is that the mind *projects* this sort of association onto events and objects in the sensible world. Necessary connections are not features of the sensible world *per se*, but are products of the mind’s experience of it. More specifically, it is simply that human beings get used to seeing patterned events, and are conditioned into thinking that events are necessarily related. In reality, however, all there is only constant conjunction between events, and the customary transition of the mind from one particular event to another. In Hume’s words again, causation:

…is something that exists in the mind, not in objects, nor is it even possible for us ever to form the most distant idea of it, considered as a quality of bodies…The efficacy or energy of cause…belongs entirely to the soul, which considers the union of two or more objects in all past instances. ‘Tis here that the real power of cause is plac’d, along with their connexion and necessity (Hume, 2000: 112).

It is important to note that, notwithstanding these arguments, Hume is a determinist: he is committed to the view that all events are caused. This view, the determinist thesis, is a pillar of the Newtonian scientific worldview to which Hume was firmly committed. Thus although Hume has his distinctive analysis of what we are doing when we make causal judgments, and of what those judgments are based on, none of that undermines his confidence in the basic determinist worldview. The argument that judgments of causal necessity are based on habit induced by exposure to constant
conjunction is not supposed to undermine our causal judgments, but to show that they have a different foundation than we thought (not in reason, but in our nature). The importance of this discussion, then, has not been to show that Hume is a skeptic generally about causation, but that he is skeptical about the objective provability of the causal relation through an appeal to reason. Bearing this in mind, let us turn our attention to Kant.

(iii) Kant on causation

Kant’s view of causation can be seen as a direct response to Hume. Kant argues in the “Second Analogy” of the first Critique that we can establish a priori that each event is determined to occur by some preceding event in accordance with a causal law.

It is important to understand Kant’s conception in light of the transcendental idealist philosophy that he develops and defends in the first Critique as well as the Prolegomena to Any Future Metaphysics, and which informs all his mature writings. This philosophy is an epistemological response to the classical traditions of rationalism and empiricism. For both traditions, the mind is represented as passive insofar as it is characterized as either (i) already in possession of innate ideas ready to be analyzed, or (ii) a “blank slate” ready to be impinged upon by external sensation. Kant’s insight is that the mind plays an active role in our experience of reality. In the “Transcendental Aesthetic” section of the first Critique and the “First Part” of the Prolegomena, Kant argues that our mind’s receptivity to the external world—our sensibility—is governed by the a priori intuitions of space and time. In other words, space and time make up the
form of our sensibility; they are innate intuitions that shape the way we perceive the
world. Kant demonstrates this through an analysis of geometry and arithmetic, showing
that the possibility of pure mathematics proves as much.

Beginning with geometry, Kant argues that previous philosophers have
mistakenly claimed that geometrical proofs are *a priori* analytic judgments.¹ Instead,
Kant claims they are *a priori* synthetic.² Their syntheticity relies on the pure intuition of
space, and this can be demonstrated by carefully analyzing geometric judgments. As
Ellington states in his introduction to Kant’s *Prolegomena*: “When one says that a
straight line is the shortest distance between two points, he makes an appeal to spatial
intuition. The concept of straight is merely qualitative. The concept of shortest is not
already contained in the concept of straight but is an addition to straight through recourse
to the pure intuition of space” (Kant, *Pr*: x-xi).

Kant also argues that time, like space, is a formal feature of our sensibility, which
can be demonstrated through an analysis of pure arithmetic and mechanics. Like
geometry, Kant claims that arithmetic and mechanics represent *a priori* synthetic
judgments, whose syntheticity rests on the addition of the *a priori* intuition of time. As
Kant puts it: “Arithmetic attains its concepts of numbers by the successive addition of
units in time, and pure mechanics especially can attain its concepts of motion only by

---

¹ By definition, this means that geometric judgments: (i) precede experience (are *a priori*), (ii) are
subject to the principle of contradiction (their denial necessarily results in one), and (iii) are
subject to conceptual containment (their judgments express nothing in the predicate but what has
been already actually thought in the concept of the subject, e.g. “All bodies are extended”) (Kant,
*Pr*: 12).
² Once again, this type of judgment precedes experience (*a priori*), but “contains in its predicate
something not actually in the universal concept”, e.g. “Some bodies have weight” (Kant, *Pr*: 12).
employing the representations of time” (Kant, Pr: 27). Kant’s principal point is that pure mathematics—which is in fact composed of a priori synthetic judgments—is possible because of our a priori intuitions of space and time. More generally, this is supposed to show that it is impossible for us to have any experience of objects that are not in time and space: they are intuitions that the mind necessarily brings to its experience of the world, and so which condition all the actual impressions that a subject receives through sensibility. This leads Kant to conclude that the sensible world is the world of appearances. The world as we experience it is essentially conditioned by our forms of intuition and thus we only know things as they appear to us, not as they really are.

In the “Analytic of Concepts” section of the Critique and the “Second Part” of the Prolegomena, Kant reminds us that experience also consists in judgments, which belong to the faculty of the understanding (as opposed to that of sensibility). Kant asserts that judging can either consist in the subjective comparison and connection of perceptions in a single subject, or in the comparison and connection of perceptions “in consciousness in general” (Kant, Pr: 43). To be able to move from the former types of judgment (“judgments of perception”) to the latter (“judgments of experience”), Kant reasons that something must be added to sensory intuitions to make them universally and objectively valid (Kant, Pr: 41-2). These additions are what Kant terms the pure concepts of the understanding. He writes: “A concept of this nature is a pure a priori concept of the understanding, which does nothing but determine for an intuition the general way in

---

3 For a full view of Kant’s arguments about space and time see the “Transcendental Aesthetic” in
which it can be used for judging” (Kant, *Pr*: 44). To demonstrate the function of the pure concepts of the understanding in the *Prolegomena*, Kant offers the following judgment: When the sun shines on the stone, it grows warm. Though this judgment is merely one of perception, we may also make the following judgment: The sun warms the stone.

According to Kant, what allows us to make this latter judgment is the addition of a pure concept of the understanding to the perception—the concept of cause—which necessarily connects the concept of sun with the concept of heat, and hence makes the judgment universally valid and objective. The pure concept of cause therefore enables us to move from subjective judgments of perception to objective judgments of experience, and so allows us to compare and connect perceptions in consciousness in general.

Kant’s point is that just like how our sensibility presupposes the features of space and time, judgments of experience are only possible by presupposing the possession of certain fundamental *a priori* concepts that the understanding brings to experience. Kant presents all of these concepts in both his *Critique* and *Prolegomena* in the form of his “Transcendental Table of the Concepts of the Understanding”. The table is comprised of four divisions (quantity, quality, relation, and modality), and the concept of cause falls under that of relation. Having now provided a general sketch of Kant’s epistemology, let us now turn our attention to causation specifically, in order to see how Kant represents causality as an *a priori* concept of the understanding, and so a necessary feature of experience generally.

the first *Critique* and the “First Part of the Main Transcendental Question” in the *Prolegomena*. 

10
In the “Second Analogy”, Kant enters the discussion of causation by firstly appealing to how we come to experience objects in terms of the distinction between what we do not and what we do take to be a causal relationship. In technical terms, this is the distinction between what he calls the “succession of appearances” and the “appearance of succession” (Kant, CPR: 304). The former Kant defines as a subjective sequence, where the succession is not informative about the object. The latter, conversely, is characterized in terms of an objective sequence, where one appearance seems to depend on that which preceded it. Kant employs an example that considers the difference between the appearance of a house, and the appearance of a ship moving downstream. Concerning the former, whatever order the parts of the house appear to the observer is entirely dependent on the observer’s subjectivity. In other words, the house can present itself to the observer in a number of different ways, and these possibilities are contained in the subject as opposed to the object: “…my perceptions could have begun at its rooftop and ended at the ground, but could also have begun below and ended above…In the series of these perceptions there was therefore no determinate order that made it necessary when I had to begin in the apprehension in order to combine the manifold empirically” (Kant, CPR: 307). Kant’s point is that, although the experiences of the house are ordered temporally, temporal succession is not a quality of that which the appearances represent to us.

In the ship example, however, the same interpretation does not hold. The way in which the ship’s movement (at one location in the sea followed by a further location) is apprehended in the observer is not a consequence of the observer’s subjective ordering of
the perceptions. Instead, the ordering seems to be contained in the object itself. In Kant’s words: “My perception of its position downstream follows the perception of its position upstream, and it is impossible that in the apprehension of this appearance the shop should first be perceived downstream and afterwards upstream. The order in the sequence of the perceptions in apprehension is therefore here determined, and the apprehension is bound to it” (Kant, CPR: 307). So contrary to the succession of appearances, with an appearance of succession, we take one event to follow another objectively, not merely in terms of appearances.

According to Kant, the distinction between the house and the ship examples leads to certain conclusions about connections within the manifold of possible experience. The house example can say nothing about the connection among objects of possible experience because as noted above, the successive appearance of the house is determined arbitrarily by the subjective observer. The ship example, however, tells us something about connections within the experienced manifold because in this case, the perceived connection between objects of possible experience is not arbitrary. This case functions in such a way that one appearance is necessarily followed by another, and it is impossible to reverse this order so that the second occurrence could have preceded the first.

Kant’s argues that we must understand this transition in terms of a causal law. His point is that the sequence of our perceptions in these sorts of cases has been determined in accordance with a rule—that states of the second type can only follow states of the first type. Furthermore, although the causal sequence is apprehended by examining our experiences of the world, Kant maintains that it is in fact grounded a
This is so because if causal laws are not presupposed, all human understanding of sequence would be determined subjectively and we would not be able to draw the distinction between the succession of appearances and the appearance of succession in the way that we do. The causal law, therefore, is a precondition for all human experience.

Thus Kant’s insight against Hume is that we should not try to derive the justification for our causal concepts by generalization from experience—Hume is right that this is a hopeless project; rather, the applicability of causal concepts is a presupposition of the very possibility of experience in the first place. As D.P Dryer puts it:

Hume took it for granted that we discover that A is the cause of B by induction from observations of A-like events being followed by B-like events and that it is thus that we are led to form the concept of cause. Hume therefore assumes that it must first be known what follows upon what before the concept of cause is reached. But what Kant points out is that it cannot be known what follows upon what without the concept of cause being assumed (Dryer, 1984, 59).

(iv) The moral law: Kant’s categorical imperative

So far we have seen how Kant argues that the world of appearances is necessarily governed by causal law. To make sense of moral action and responsibility, however, Kant must somehow incorporate freedom into this causal schema. Let us therefore consider Kant’s account of freedom, as presented in the *Groundwork* and the *Second Critique*. I begin with some preliminaries about Kantian moral philosophy, drawing specifically on Chapters I and II of the *Groundwork*. 
Kant’s task in the *Groundwork* is to establish *a priori* the supreme principle of morality, a principle that can hold unconditionally for all rational agents. He begins by examining ordinary moral judgments, and concludes that the only thing good without qualification is a good will (Kant, *Gr*: 18). All other seemingly good qualities, like courage and honor, may be used for ill purposes in certain circumstances. They are not intrinsically good—instead, their goodness depends on something else, namely, the prior possession of a good will. So, if a good will is the only unconditioned good, what is it that makes it good? It cannot be, Kant argues, the results it produces, for the goodness or badness of whatever results depends on, in the first place, the presence of a good will. Results are thus conditioned by the will’s goodness or badness; they cannot be the source of the will’s unconditioned goodness. Further, a good will retains its goodness whether or not the desired results are achieved.

Kant concludes that the goodness of the will must be based on what motivates the will’s willing. But what specifically should motivate the will? Kant claims firstly that the will’s goodness cannot be attained through the motives of self-interest or inclination. These are subjective and accidental motives, which cannot form the basis of an objective *a priori* morality. An action only has moral worth, claims Kant, if it is determined in accordance with duty, which he defines as reverence, or respect for the moral law (Kant, *Gr*: 21). A good will, therefore, is one whose will is determined in accordance with the moral law.

What, however, is the moral law? Laws, generally, are those things that apply to us in virtue of being members of a group, be it a societal group, a political organization,
or a sports team. They provide us with guidelines and directives for how we ought to, and ought not to act as members of the group in question. The moral law, for Kant, is that principle which applies to us in virtue of being a rational member of humanity. The moral law is thus to be derived through reason, and applies universally to all rational members of society.\(^4\) It is therefore the underlying principle guiding \textit{all} behavior—not just behavior pertaining to a particular societal group—insofar as we are all rational beings, we are members of the group for which the moral law applies unconditionally.\(^5\)

Thus far we have looked at the source and structure of the moral law. We must now consider its content. Kant calls the moral law the categorical imperative. It is categorical insofar as it binds us unconditionally (as opposed to it making reference to ends, as does the “if-then” of the hypothetical imperative); it is an imperative insofar as it commands us. Though Kant presents several formulations of the categorical imperative in the \textit{Groundwork}, the first and most important one is known as the \textit{Formula of the Universal Law}. It states the following: “I ought never to act except in such a way that I can also will that my maxim should become a universal law” (Kant, \textit{Gr}: 22). Though seemingly contentless, this first formulation commands us only to will actions that we could rationally want any human being to will. I, as an acting agent, take the way in which I would like to act in whatever particular circumstance, and see whether or not this

\(^4\) See Johnson, 2002.
\(^5\) Certainly it is not the case that all humans are necessarily rational (children, and the psychologically abnormal cannot readily be seen as such). The point, however, is that reason is something that the vast majority of mature adults are equipped with, and insofar as we possess this capacity, the moral law applies for us. It is thus not like any other sort of membership, where we may legitimately opt out, and therefore no longer be subject to the laws of that group. Unlike
sort of action could be rationally willed for all humans in such circumstances. For Kant, actions that turn out to be immoral are those which we cannot will universally because (i) they pose a logical contradiction, or (ii) it would otherwise be irrational to universalize the action (for example, if it would be self-defeating to suppose that others should act as I am proposing to act). Throughout Chapters II and III Kant proposes various different formulations of the categorical imperative, though he also claims that they are equivalent, representing different expressions of the same law. These formulations are known in the literature as the Respect for Persons Formula, Principle of Autonomy Formula, and Kingdom of Ends Formula. For the purposes of this chapter, I will not provide a discussion of these various formulations. What is critical is that if agents are to act out of respect for the moral law, they must be genuinely free: their action must be wholly determined by the rational will and free of the influence of inclination, desire, and other causal factors. Having provided a sketch of Kantian morality, I turn now to his arguments for freedom—the necessary condition for legitimating his ethics.

(v) Kant’s arguments for freedom

In this section I address two of Kant’s arguments about freedom, as presented in Chapter III of the Groundwork and the second Critique.

---

these sorts of laws, the moral law applies to us in virtue of simply being a rational being—a status from which one cannot legitimately opt out.

6 For a full discussion of the Categorical Imperative and its various formulations, see Chapter II of Kant’s Groundwork. For an analysis/summary of these same arguments, see Paton’s The Moral Law, pgs. 24-38.
In Chapter II of the *Groundwork* Kant establishes the *Principle of Autonomy*, which, as we noted earlier, is one of the various formulations of the categorical imperative. The principle states: “So act that your will can regard itself at the same time as making universal law through its maxim” (Kant, *Gr*: 33). It adds to the earlier formulation in that it prescribes not only that we act in accordance with laws that can be universalized, but also that these laws be self-legislated through the rational will. Clearly, such self-legislation presupposes that the will is free. Though Kant here shows the reciprocal relationship between the moral law and freedom, he reminds us that he has yet to provide any grounding for either concept: “We have merely shown by developing the concept of morality in vogue that autonomy of the will is unavoidably bound up with it or rather at its very basis” (Kant, *Gr*: 106).

In Chapter III, Kant sets out to ground these concepts. He begins by defining freedom as the property of the rational human will, which is somehow aloof from the causal system of the physical world. Kant asserts that freedom must refer to the will’s ability to operate through self-imposed laws. This definition reflects the earlier *Principle of Autonomy*, and thus Kant points out that presupposing freedom leads necessarily to the *Principle*, and therefore to the possibility of morality. This analysis, however, takes us no further, merely reasserting the reciprocal relationship between freedom and the moral law.

In what is often interpreted as an odd turn in Kant’s argument, he next puts forth what is known as the “preparatory argument”. Kant tries to show that freedom must be presupposed as a property of the will of all rational beings. Kant believes he can
demonstrate this by appealing to the fact that all rational beings always act under the Idea of freedom. This means, simply, that the form of human thought always presupposes freedom—we see ourselves as the sole authors of our thoughts and decisions, not as parts of a greater causal system.

Upon fleshing out the “preparatory argument” Kant acknowledges that he has yet to make any progress in terms of grounding the moral law or freedom. Indeed, it seems his argument has now become circular. First Kant showed by analytic argument that the Principle of Autonomy led necessarily to the Idea of freedom. Next, he argued that the Moral Law applies to us because we always act under the Idea of freedom. As Kant puts it: “…in the order of efficient causes we take ourselves to be free so that we may conceive of ourselves to be under moral laws in the order of ends; and we then proceed to think of ourselves as subject to moral laws on the ground that we have described our will as free” (Kant, Gr: 111).

The solution to this circularity lies in an important tenet of Kant’s transcendental idealism, namely, the distinction he draws between the phenomenal and noumenal realms. Recall that the phenomenal world represents the sensible world of appearances, wherein experience is conditioned by the a priori forms of our intuition and the concepts of the understanding. This recognition, however, of the sensible world’s conditioning, naturally invites us to seek the unconditioned. This is merely that which lies beyond appearances: thing-in-themselves, or noumena. Kant’s claim, in terms of morality, is that we may remove the above-noted vicious circularity by viewing ourselves as members of both the phenomenal and the noumenal worlds. From one standpoint, we may regard
ourselves as members of the latter: as passive, causally conditioned members of the sensible world. We may also regard ourselves, however, as members of the former realm: as free agents existing in an intelligible realm that goes beyond the sensible world of experience. We are entitled to envisage ourselves as members of the noumenal realm, Kant asserts, because we possess the capacity of reason, a faculty that is by definition completely removed from the sensible world. As Allison writes in his book, *Kant’s Theory of Freedom*, we can demonstrate our possession of this faculty because: “…in the consciousness of our epistemic spontaneity we are directly aware of a capacity we cannot conceive as sensibly conditioned” (Allison, 1990: 222). Kant claims that since reason operates independently of natural necessity, the laws operating within this faculty must be grounded in the Idea of freedom, which is conceived simply as the ability to act outside of the phenomenal world. Thus Kant grounds our entitlement to posit ourselves as members of an intelligible world in the faculty of reason, which is in turn grounded in the Idea of freedom. Having therefore grounded freedom without reference to the moral law (and already established the reciprocity thesis between freedom and morality), Kant is now able to secure the bindingness of the categorical imperative in a linear fashion: reason, freedom, morality.

Before providing a more thorough and critical analysis of the noumenal/phenomenal distinction, let us briefly turn our attention to Kant’s second argument for freedom, as presented in the second *Critique*. Unlike the argument from the *Groundwork*, the argument in the second *Critique* explicitly denies the possibility of any deduction of the moral law and claims instead that the moral law—understood as a “fact
of reason”—can serve as the basis for a deduction of freedom. Although the texts are far from unambiguous on this score, the bulk of evidence suggests that the fact of reason is best construed as the consciousness of standing under the moral law. In other words, in our experience we always deem ourselves to be operating within the consciousness of moral constraint. In Stephen Engstrom’s introduction to Kant’s second Critique, he explains the fact of reason by noting the etymology of the word: “fact”, he tells us, comes from the Latin translation of “factum”, meaning deed or action (Kant, CPrR: xlii). He goes on to characterize the fact, or deed of reason in the following manner:

Unlike the familiar deeds that human beings perform in practical life, which are occurrences that take place in time, a deed of pure reason is not an action undertaken in response to any specific empirical conditions and cannot be assigned to any particular position in time; it is rather the activity of reason itself, manifest in practical life as a fundamental law, something unchangingly operative at all time and in all conditions, even if its effect may vary owing to differences in the conditions and circumstances in which it is operative (Kant, CPrR: xlii).

Kant’s idea seems to be that the fact of reason, understood as a kind of pure law that governs practical experience, is synonymous with the moral law—we can thus deduce the moral law from the fact of reason. In the Corollary to §7 in the second Critique, Kant states: “Pure reason is practical by itself alone and gives (to the human being) a universal law, which we call the moral law” (Kant, CPrR: 46). From here, the move to freedom becomes clear due to the reciprocity thesis. In Kant’s own words: “Therefore is it the moral law of which we become conscious directly…which first offers itself to us, and which…leads straight to the concept of freedom” (Kant, CPrR: 42). It is important to recognize that freedom, here described, has received practical, though not
theoretical justification. Kant thus terms it the first postulate of pure practical reason, which he defines as: “…a theoretical proposition, though one not provable as such, insofar as it attaches inseparably to a practical law that holds a priori and unconditionally” (Kant, CPrR: xliv). Recognition of the moral law, derived through the fact of reason, furnishes practical reality to freedom.

We can now see how the phenomenal/noumenal distinction fits once again into this second argument for freedom. It is our consciousness of the moral law (and with this our practical recognition of freedom) that provides evidence of our existence in an intelligible world, removed from the causal laws governing the phenomenal world. In contrast to the argument put forth in the Groundwork, where our possession of the capacity of reason led Kant to suppose our existence in an intelligible realm, he posits this same realm quite differently in the second Critique. In the latter argument, his inference to the intelligible world springs from our initial consciousness of the moral law.

(vi) Kant’s noumenal/phenomenal distinction: an explanation and a criticism

We have now surveyed Kant’s two arguments for freedom, and though they are different in important respects, they both end up locating freedom in the realm of the noumenal. Let us now take a closer look at the noumenal/phenomenal distinction, and address certain pertinent problems with this noumenal conception of freedom. The phenomenal world, for Kant, is the world of sensible experience. In this world, objects affect us through our senses. These objects exist independently of us, in the sense that they do not depend upon us causally for their existence. Recall, however, that our experience of
these objects is transcendentally ideal, since as Kant has argued in the first *Critique*, our experience is constituted through our *a priori* intuitions of space and time, as well as the pure concepts of the understanding (the categories). Because of this, Kant concludes that we can only ever come to know objects *as they appear to us*, not as they really are.

This recognition naturally invites us to posit the noumenal realm. Stated by Kant in the first *Critique*:

…if we call certain objects, as appearances, beings of sense (*phaemonema*), because we distinguish the way in which we intuit them from their constitution in itself, then it already follows from our concept that to these we as it were oppose, as objects thought merely through the understanding, either other objects conceived in accordance with the latter constitution, even though we do not intuit it in them, or else other possible things, which are not objects of our senses at all, and call these being of understanding (*noumena*) (Kant, *CPR*: 360).

Here the central point is analytic: “it is a feature of the *concept* of appearance that to apply that concept to some existent is to imply that there is some existent to which applies the concept of the in itself—the concept of a ground of appearance that is in some way independent of the appearance” (Franks, 2005: 44). Noumena, therefore, represent what Kant calls the *thing-in-itself*: whatever exists beyond the bounds of sense. The concept of noumenon is therefore only a negative concept, what Kant sometimes refers to as a “boundary concept” (Kant, *CPR*: 362). It is a concept that we can conceive of, but never have knowledge of. Knowledge of a thing-in-itself is not possible because it represents merely that which is completely independent of our modes of access to it (the modes of space and time, and the pure concepts of the understanding). As human beings, we can only come to know things through these modes and so, as appearances. Kant
writes: “The concept of a noumenon is therefore merely a boundary concept…But it is
nevertheless not invented arbitrarily, but is rather connected with the limitation of
sensibility, yet without being able to posit anything positive outside of the domain of the
latter” (Kant, CPR: 362).

Returning to Kant’s ethics, we can now see more clearly how the
noumenal/phenomenal distinction functions to incorporate freedom into his overall
picture. As we have seen, the sensible world is governed by the a priori intuitions of
space and time and the pure concepts of the understanding. For Kant the concept of
causality is necessary for a coherent and unified experience of objects. Freedom,
therefore, is incompatible with Kant’s depiction of the phenomenal world, since our
experience of the sensible world is exhaustively governed by causal principles. Kant’s
transcendental idealism, however, allows him to invoke the distinction between noumena
and phenomena, and to locate freedom in this former realm—that realm which is neither
subject to the formal intuitions of space and time nor to the pure concepts of the
understanding.

Kant’s strategy, however, provokes an obvious objection. It is tempting to
suppose that when Kant draws the distinction between the phenomenal world and the
noumenal world, he is actually making a distinction between two separate and unique
realms. Paul Franks, in his book All or Nothing, terms this interpretation “the two
existents interpretation”, and characterizes it in the following way: “On this view, Kant
speaks of things in themselves referentially, and he is committed to the existence of
entities distinct from the sensible objects of our knowledge. These entities inhabit the
intelligible world, which is distinct from the sensible world” (Franks, 2005: 37). On this two-worlds, or two-existents interpretation, freedom is thus conceived of as an entity ‘inhabiting’ the noumenal world, where this world is completely distinct and separate from the phenomenal world. The problem here is obvious: if freedom exists in a separate and unique world, how do we make sense of freedom as something that we, as inhabitants of the phenomenal world, possess? The sharp two-world distinction that this interpretation holds to makes it difficult to see just how noumenal freedom can affect our lives as sensible beings. Furthermore, even if we grant that freedom can somehow make a difference in the sensible world, this would destroy both the timeless status of the noumenal world and the causal principle governing the phenomenal world.

A further difficulty derives from Kant’s general characterization of the noumenal realm as impossible to characterize in contentful thought, thus making freedom, as a constituent of this realm, effectively unknowable. As stated by Bernard Carnois in his book *The Coherence of Kant’s Doctrine of Freedom*:

> Freedom in itself is not merely unknowable, it is also unexplainable. The knowledge of freedom remains limited precisely because the objective reality of the concept ‘can in now way be shown according to natural laws or in any possible experience’. For all possible explanation ends where determination by natural laws ends. We can and must admit that freedom is possible, but we cannot explain *in what way* it is possible (Carnois, 1987: 71).

If freedom is noumenal, then its status and its influence must remain essentially obscure. As Kant himself puts it in the *Groundwork*: “Freedom…is a mere Idea: its objective validity can in no way be exhibited…Thus the Idea of freedom can never admit of full comprehension, or indeed of insight…” (Kant, *Gr*: 119).
There is, however, a more plausible interpretation of the noumenal/phenomenal distinction available. It is open to us to hold that the phenomenal and noumenal realms are not two distinct worlds, but two different perspectives for thinking about ourselves—at once as sentient beings subject to causal law, and as rational beings capable of free action. In the third chapter of the *Groundwork* Kant writes “…a rational being must regard himself *qua intelligence*…as belonging to the intelligible world, not to the sensible one. He has therefore *two points of view* from which he can regard himself and from which he can know laws governing the employment of his powers and consequently governing all his actions” (Kant, *Gr*: 113). Echoing this same two-perspective sentiment in an equally telling passage, he claims: “We see now that when we think of ourselves as free, we transfer ourselves into the intelligible world as members and recognize the autonomy of the will together with its consequence—morality; whereas when we think of ourselves as under obligation, we look upon ourselves as belonging to the sensible world…” (Kant, *Gr*: 113). On this view, it is not that the phenomenal world and the noumenal world are ontological concepts—distinct places, as it were. It is rather that human agents are open to two distinct modes of interpretation: we can see ourselves as inhabitants of the natural order, and when we do our behavior is fully explicable by appeal to the causal principles that govern natural events. But we can also see ourselves as autonomous *rational* beings, who are able to act in light of rational principles by which they govern their own activity.

If we take the two standpoints interpretation, we need no longer be concerned with the issue of how freedom, situated somehow in a timeless realm beyond the natural
order, can still somehow intervene in nature. Rather, the question becomes: what is the relation between the two standpoints? How can it both be true that we are properly understood as natural beings and as rational agents? How can the two standpoints both be true of the same thing? Kant himself does not really address these questions in these terms, for he remains too wedded to the idea that freedom is a substantial presence. But the two standpoints approach can be seen resurfacing in the work of many thinkers inspired by Kant, and in the rest of this thesis we shall explore how two of them, McDowell and Strawson, deploy something like it in their respective approaches to freedom.

7 In this interpretation, I follow Henry Allison’s view in his book Kant’s Theory of Freedom. Allison writes, “it is relatively uncontroversial that at the heart of Kant’s account of freedom in all three Critiques and in his major writings on moral philosophy is ... an explicitly indeterminist or incompatibilist conception (requiring an independence of determination by all antecedent causes in the phenomenal world)” (Allison, 1990: 1; see also pp.47-8 where Allison discusses Kant’s view of reason as a “timeless causal power”).
Chapter 2

(i) McDowell’s overall project in *Mind and World*

Thus far I have considered Kant’s conceptions of causation and freedom and argued that his view of their relation, with its appeal to the noumenal/phenomenal distinction, is philosophically unsatisfying. In Chapter Two, I argue that McDowell’s view of the relation between nature and “the space of reasons”, developed in his book *Mind and World*, provides a persuasive alternative to the Kantian picture. In Kantian terms, McDowell’s space of reasons is analogous to the realm of spontaneity, where the understanding has the capacity to form thoughts and make judgments—and thus operates freely. As McDowell puts it: “When Kant describes the understanding as a faculty of spontaneity, that reflects his view of the relation between reason and freedom: rational necessitation is not just compatible with freedom but constitutive of it. In a slogan, the space of reasons is the realm of freedom” (McDowell, 1994: 5). Before exploring McDowell’s account of the space of reasons as an alternative to Kant’s noumenal conception of freedom, I begin by presenting McDowell’s overall project in his seminal text.

McDowell begins *Mind and World* by suggesting that the history of modern philosophy has faced an enduring difficulty in finding a satisfying account of the relation between thought and reality. The difficulty stems from the fact that the life of the mind contains two separate and distinct features. There are the deliverances of experience in
which the physical, natural world (“external reality”) impinges upon us in sensation. There is also conceptualized thought, composed in part of judgments, or beliefs, about the world.\textsuperscript{8} In light of this distinction, the challenge facing epistemologists is to show that our thoughts are rationally related to our sensible experiences of the physical world. In other words, the task is to show that our judgments and beliefs, in relation to external reality, can somehow be justified—and therefore count as knowledge. As McDowell puts it in the opening passages of *Mind and World*: “…surely there must be such grounding if experience is to be a source of knowledge, and more generally, if the bearing of empirical judgments on reality is to be intelligibly in place in our picture at all” (McDowell, 1994: 5).

According to McDowell, solutions to this epistemological challenge have historically appealed to one of two alternatives, which he finds to be equally problematic: (i) the idea of the Given, and (ii) Coherentism. The former holds that relations of justification between mind and world are grounded in unconceptualized sense data. The Given is something that can be known directly—an immediate epistemic episode—and which can therefore serve as a foundation for higher conceptualized knowledge. Richard Bernstein writes: “Although given-ness has taken on many different forms in the history of philosophy, the Given is presumably not ‘contaminated’ by any mediation, any form of inference. The Given is something that can be either directly known or known by acquaintance; it is an immediate intuition or self-authenticating episode” (Bernstein, 2002: 11).

\textsuperscript{8} In Kantian terminology this is the distinction between “receptivity” and “spontaneity”.

28
McDowell argues, following Sellars, that the Given, thus understood, could provide a causal, but not rational (and thus justificatory) influence upon judgment and belief. If we appeal to non-conceptual sense data to ground conceptualized thought, this appeals to something that exists outside of the conceptual realm. Understanding thought as connected to the empirical world in this way demonstrates merely a causal link between thought and reality. As Barry Stroud puts it:

…to understand two things as connected in accordance with a law of nature is not to understand one of those things as making the other reasonable in any way, or as justifying or supporting it. It is to see that one thing happens because something else happens, but not to see that one thing is correctly or incorrectly executed, or is warranted or justified in the light of something else (Stroud, 2002: 80)

Using McDowell’s own terminology to express this same point, the problem with the Given is that “the space of reasons”—the “space of justifications and warrants”—is made to extend past the conceptual sphere, towards this un-conceptualized given-ness: “What we wanted was a reassurance that when we use our concepts in judgment—our spontaneity…is constrained from outside thought, and constrained in a way that we can appeal to in displaying the judgments as justified” (McDowell, 1994: 8). However, the resulting picture looks like one in which conceptualized thought is related to the world through “a brute impact from the exterior” (McDowell, 1994: 8). We end up with a picture in which the world is seen as an “alien force”, causally impacting us.

In his footnote in Lecture I (pg. 8), McDowell likens the picture offered by the Given to a person caught in a tornado. McDowell invites us to imagine a situation wherein the sheer force of the tornado deposits the person in a land they were restricted
from entering. Though it is clear that the empirical world in this case impacted the person physically, the result of this impact is in no sense rationally connected to the person’s spontaneity. In fact, we would assert that this person should be ‘exculpated’ from responsibility for what happened insofar as she was powerless to exert any control over this force. According to McDowell, conceiving of the mind-to-world justification by appealing to what he pejoratively calls “the Myth of the Given” results in an analogous state of affairs: it can explain how we are impacted from the exterior in a causal sense, but it offers nothing in terms of the rational relations holding between mind and world. Stated by McDowell “…it is one thing to be exempt from blame, on the ground that the position we find ourselves in can be traced ultimately to brute force; it is quite another thing to have a justification. In effect, the idea of the Given offers exculpations where we wanted justifications” (McDowell, 1994: 8).

The alternative picture is Coherentism. This option, mindful of the weakness of the Myth of the Given, rejects the idea that the world can itself provide any rational justification for our thought. The main thrust of Coherentism then is to suggest that what justifies our possession of whatever beliefs are simply other beliefs. To quote Davidson, whom McDowell treats as the principal representative of this alternative: “Nothing can count as a reason for holding a belief except another belief” (McDowell, 1994: 14). On this view, then, a belief is justified if and only if it coheres with the overall belief-set. Though the world certainly plays a role in the Coherentist picture, it plays merely a causal role. The rational justification for our beliefs has no external constraint—
justification relies solely on the presence or absence of other beliefs in the overall belief-set.

There is an obvious problem with this alternative, which McDowell is quick to point out. If our goal is to make sense of how our thought comes to bear on external reality, it looks as if some external constraint regarding justification must be met, but this Coherentism does not provide. Instead it claims that justification simply does not require such a constraint. Coherentism, therefore, leaves us with the picture of a free-floating conceptual realm that is not tied down to the real world. As Stroud writes: “…with no experiential ‘grounding’ thought will not be answerable to anything independent of thought, and so will degenerate into a series of ‘moves in a self-contained game (p.5), ‘a frictionless spinning in a void’ (p.11)” (Stroud, 2002: 80). In McDowell’s own words: “It can seem that [in appealing to Coherentism] we are retaining a role for spontaneity but refusing to acknowledge any role for receptivity, and that is intolerable. If our activity in empirical thought and judgment is to be recognizable as bearing on reality at all, there must be external constraint” (McDowell, 1994: 15). We are thus driven back to the idea of the Given as the only thing that can provide the necessary friction with reality.

McDowell’s proposal, in Mind and World, is to suggest a way out of the oscillation between the Myth of the Given and Coherentism. Deploying a Kantian idiom, McDowell asserts that we must rethink the underlying understanding of the nature of “receptivity” and “spontaneity” that both the Myth of the Given and Coherentism presume. Recall that receptivity refers to the external world’s impingement onto us through sensation, whereas spontaneity refers to conceptualized thought, including
judgments and beliefs. According to McDowell, both theories try to locate justificatory relations between mind and world under the presumption that there is a strict separation between these two realms. More specifically, the presumption is that the world impinges onto us through un-conceptualized sense data, whereas the mind’s judgments and beliefs take place in a separate, conceptualized realm. Receptivity and spontaneity, therefore, are sharply separate and fundamentally different in nature. McDowell suggests that we return to the Kantian insight which sees experience as consisting in a combination of receptivity and spontaneity. In experience, receptivity should not be understood in terms of the impingement of un-conceptualized sense data; and spontaneity should not be treated as a conceptual realm at remove from the externality of receptivity. There is neither room for un-conceptualized sensory input standing in no relation to conceptual thought, nor pure thought operating independently of all rational constraint from our sensible experiences.

Instead, we should understand that conceptual capacities are “drawn on in receptivity” (McDowell, 1994: 9). In other words, our experience in terms of what we receive in sensation is not strictly non-conceptual, as both the Myth of the Given and Coherentism presume, but already equipped with conceptual capacities. As Robert Pippin puts it:

…conceptual capacities are to be “operative” not only in judgments, but “already in the transactions in nature that are constituted by the world’s impacts on the receptive capacities of a suitable subject” (pg. xx). This is to be distinguished from the claim that conceptual capacities are exercised “on non-conceptual deliverances of sensibility (pg. 39). For McDowell, by contrast, such deliverances are already conceptual, or “capacities that belong to spontaneity are in play in actualizations of receptivity (Pippin: citing McDowell, 1994)
According to McDowell, characterizing experience in this way—recognizing the unboundedness of the conceptual—resolves the epistemological challenge. With this framework in play we first of all retain the feature of external control which the Myth of the Given preserves and Coherentism dismisses, by retaining a role for receptivity. Stated by McDowell: “…experiences are indeed receptivity in operation; so they can satisfy the need for an external control on our freedom in empirical thinking” (McDowell, 1994: 24).

Moreover, we can now demonstrate how receptivity can serve as a justificatory role for beliefs and judgments, and not just a causal one (which as we saw earlier was the problem facing the Myth of the Given). This can be achieved since unlike with the Myth of the Given, experiences at the level of receptivity are already equipped with conceptual content. In McDowell’s own words, what a person takes in in an experience is: “that things are thus and so” (McDowell, 1994: 25). This does not mean, however, that sense experience is just another form of belief of judgment. The world’s impact on us is certainly an articulation of our receptivity as opposed to our spontaneity, insofar as it is passive as opposed to active. So in sensation the external world impacts us independently of our control. However, because the passive impact of the world in receptivity is already conceptual, there is no ontological gap between what one takes to be thus and so in receptivity and what one takes to be thus and so in judgment. In this sense our experience of the empirical world, which is passive but already conceptual, can serve as a rational constraint on our spontaneity. Stated by McDowell: “That things are
thus and so is the conceptual content of an experience, but if the subject of the experience is not misled, that very same thing, that things are thus and so, is also a perceptible fact, an aspect of the perceptible world” (McDowell, 1994: 26).

The point is that in McDowell’s schema, though we can still be misled by our senses (and so capable of possessing false judgments or beliefs), spontaneity is no longer necessarily confined within a self-enclosed conceptual sphere, wherein it becomes impossible to show a rational connection between this sphere and the empirical world. For McDowell, once we recognize the unboundedness of the conceptual, we can dissolve the philosophical anxieties that gave rise to the Myth of the Given and Coherentism in the first place. By maintaining the distinction between receptivity and spontaneity but proposing that perception is conceptual all the way down, McDowell is able to (i) conserve a role for the empirical world which Coherentism boldly dismisses, and (ii) make sense of the empirical world as serving as a rational constraint on judgment which the Myth of the Given is unable to provide.

(ii) Reason and nature: McDowell’s naturalized platonism

According to McDowell, the failure on the part of past philosophers to recognize his proposed alternative—that which sees experience as conceptual all the way down—as a viable option has its roots in a pervasive misunderstanding of sensible nature. This misunderstanding, McDowell claims, has led not only to the undesirable oscillation between the Myth of the Given and Coherentism in epistemology. It is also the underlying structure that forces Kant to adopt the noumenal/phenomenal distinction in his
quest to make sense of the relationship between nature and freedom. So although McDowell takes Kant’s insight regarding the combination of receptivity and spontaneity in experience to be correct, he disapproves of the transcendental framework in which Kant places them, which, he thinks, forces Kant to think of rationality as operating outside nature. According to McDowell, we can preserve many of Kant’s insights about rationality only if we purge them of this transcendental framework and introduce a broader conception of the natural world than Kant could countenance.

For McDowell, the conception of nature, assumed by the Myth of the Given, Coherentism, and Kant, can be traced back to the profound scientific developments of the 16th and 17th centuries. What emerged from the scientific revolution was an understanding of nature as a vast mechanism, a realm of causal law, bereft of meaning. Following Max Weber, McDowell terms this the view of nature as “disenchanted” (McDowell, 1994: 85). Characterizing sensible nature in this way led to a rigid dichotomy between nature and reason. In McDowellian terminology, the developments of the 16th and 17th centuries birthed a contrast between two types of intelligibility: “...between that sought by natural science, and the kind we find in something when we place it in relation to other occupants of the logical space of reasons—intelligibility that is proper to meaning” (McDowell, 1994: 70). According to McDowell, the strict division between natural intelligibility and intelligibility belonging (in Kantian terms) to spontaneity explains why philosophers find it so hard to see how the space of reasons—the domain of freedom—could ever somehow coexist with a law-like system of nature.
McDowell contends that Kant was in the grip of this same view of nature and this prevented him from seeing how freedom could ever be reconciled with the sensible world. This left Kant believing he had no other alternative than to constitute reason and freedom as strictly separated from the realm of nature, thus the introduction of the noumenal. As Bernstein puts it: “The sharp dichotomy that Kant makes between freedom and nature (and all the aporiai that result from this dichotomy) follow from Kant’s uncritical acceptance of such a disenchanted concept of nature” (Bernstein, 2002: 17). In McDowell’s words: “For Kant, nature is the realm of law and therefore devoid of meaning. And given such a conception of nature, genuine spontaneity cannot figure in descriptions of actualizations of natural powers as such” (McDowell, 1994: 97). Kant felt as though he had no other choice in the matter: he could either attempt to locate freedom in the realm of nature—which insofar as he conceived of this realm would effectively destroy freedom—or he could place it outside nature, and this he did.

In *Mind and World*, McDowell characterizes two styles of philosophizing that are defined by this same dichotomy: “bald naturalism” and “rampant platonism”. The former seeks to locate the space of reasons in nature through some kind of philosophical reduction or construction. As Michael Friedman puts it, bald naturalism entails “a reconstruction of reason from within the domain of nature as the object of modern natural science” (Friedman, 2002: 28). Such a strategy is obviously anathema to Kant.

The other option, rampant platonism, locates the space of reasons outside of nature, as a kind of autonomous structure. This strategy is surely reminiscent of Kant’s introduction of the noumenal realm. However, as I argued in Chapter One, this strategy
poses the threat of supernaturalism: it is unclear how we should understand ourselves as existing in both the natural realm and the removed space of reasons. McDowell contends that rampant platonism:

…pictures the space of reasons as an autonomous structure—autonomous in that it is constituted independently of anything specifically human, since what is specifically human is surely natural, and we are refusing to naturalize the requirements of reason…But human minds must somehow be able to latch on to this inhuman structure. So it looks as if we are picturing human beings as partly in nature and partly outside of it (McDowell, 1994: 77; italics mine).

Kant further intensifies this problem with his insistence that we can never have any knowledge of the noumenal realm. Recall that for Kant the physical world represents the world of appearances, where external objects appear to us in a particular way due to our a priori intuitions of space and time as well as the application of the pure concepts of the understanding. Freedom, for Kant, is located in his proposed supersensible realm, where we presume the existence of things that go beyond possible experience. These are termed by Kant as things-in-themselves, or noumena. However, since things-in-themselves are by definition independent of our experience of them, we can be nothing but utterly ignorant of them. In this sense freedom is simply a postulate, but not an object of knowledge for Kant. We are left with a dual-world schema, with little hope of understanding this dichotomy due to the fundamentally unknowable nature of the noumenal realm.

McDowell’s proposal, in Lecture IV of Mind and World is that Kant’s problematic distinction can be overcome by rethinking the conception of nature. By overcoming this disenchanted view of nature so popularized by the scientific
enlightenment of the 16\textsuperscript{th} and 17\textsuperscript{th} centuries, McDowell contends that the Kantian framework can be re-appropriated so that we can preserve the space of reasons but at the same time make sense of how it can coherently be part of the natural world. We can sustain the Kantian idea that rationality operates freely in its own sphere (but without appealing to anything supersensible), but only if we embrace a partially re-enchanted view of nature.

This re-enchantment of nature, like that conception developed out of the scientific revolution, begins by asserting that human beings are perceiving animals existing in the physical world. Humans therefore have a foothold in the sort of nature conceived of as the realm of law. However, human beings should also be defined by their \textit{second nature}, an idea that McDowell claims has been forgotten or ignored by modern naturalists. His discussion of second nature begins with the thought that human beings acquire the capacity to respond to rational requirements—to inhabit the space of reasons—as part of their normal maturation and education. For McDowell the rational requirements are there whether we are aware of them or not. Through the normal process of maturation, as well as socially structured training and practice—what McDowell calls \textit{Bildung}—we become aware of these rational requirements.\textsuperscript{9} In other words, through \textit{Bildung} our eyes become open to rational demands, we are initiated into the space of reasons (or in Kantian terms, we are initiated into the ability to operate through reason, and thus count as free). This,

\textsuperscript{9} In Rudiger Bubner’s article from \textit{Reading McDowell}, he characterizes \textit{Bildung} in the following way: “\textit{Bildung} takes place in the upbringing and civilizing of the subject, who must emerge from a state of being driven by instinct with intelligible and recognizable forms of social behavior…\textit{Bildung} means the discovery of possibilities and capacities whereby character is
in effect, is our second nature: our ability to be aware of and act on the rational requirements, to operate within the space of reasons. Our Bildung brings out our second nature, by actualizing some of the potentialities we are born with, by opening our eyes to them through a sort of education. McDowell’s inspiration for this position, which he terms naturalized or relaxed Platonism, is Aristotelian ethics, wherein a normal adult human attains what Aristotle calls “practical wisdom” through a socially educative process.

Through this appeal to a second nature, McDowell takes himself to undermine the need for the extreme Kantian distinction between noumenal and phenomenal, while avoiding the bald naturalism that no Kantian could accept. For McDowell, it is critical that we work with a conception of nature that is broader than the disenchanted conception that portrays nature exhaustively as a realm of causal law. The rational capacities we acquire in the course of our maturation and education are genuinely natural in kind, although they cannot be made intelligible by appeal to the kind of causal explanation deployed by the natural sciences. They resist exhaustive characterization in physical or biological terms but they are natural for all that. Now, McDowell maintains that our freedom consists precisely in our responsiveness to reasons. He identifies the space of reasons and the realm of freedom. But if our responsiveness to reasons can be seen as part of our natural capacities, albeit capacities that are culturally mediated and inexplicable in purely natural-scientific terms, then freedom is located within nature and is not dismissed into a supersensible realm. As McDowell writes:

shaped, not only in the direction of a socially fixed and pre-given ideal of virtue, but in the
We need to recapture the Aristotelian idea that a normal mature human being is a rational animal, but without losing the Kantian idea that rationality operates freely in its own sphere. The Kantian idea is reflected in the contrast between the organization of the space of reasons and the structure of the realm of natural law. Modern naturalism is forgetful of second nature; if we try to preserve the Kantian thought that reason is autonomous within the framework of that kind of naturalism, we disconnect our rationality from our animal being, which is what gives us our foothold in nature. We need to see ourselves as animals whose natural being is permeated with rationality, even though rationality is appropriately conceived in Kantian terms (McDowell, 1994: 85).

McDowell’s point therefore is that Kant almost got it right. What was not available to him, due to the pervasive scientism of his era, was the idea of a second nature—a way to conceive of rationality in naturalistic terms. Thus for Kant, the noumenal/phenomenal distinction looked to him like the only way to make sure that rationality was not subsumed under nature conceived as the realm of law. McDowell’s insight here is to say that if we take seriously the idea of a second nature, we can make room for the space of reasons in nature, whilst at the same time not equating the space of reasons with the realm of law. His appeal to a second nature therefore avoids the obvious Kantian problem, while still appealing to a broadly Kantian (though naturalistic) space of reasons, and thus freedom.

(iii) McDowell’s naturalized platonism: a criticism

Thus far I have argued that McDowell’s naturalistic reconciliation of freedom and nature avoids the undesirable consequences of Kant’s appeal to the noumenal/phenomenal acquisition of a personality” (Bubner, 2002: 211).
distinction; namely, making the account of freedom highly speculative. McDowell’s naturalized platonism, while upholding the reality of rational requirements, locates the space of reasons in nature, thereby avoiding an unintelligible appeal to the supersensible to ground freedom and thus make sense of moral responsibility. Remaining within a broadly Kantian framework, however, McDowell’s naturalized platonism retains a problematic feature of the Kantian picture. McDowell’s view, though it rejects the problematic distinction between the phenomenal and the noumenal nonetheless works with a dualism between what he calls the “space of reasons” and the “realm of law”. This is not quite the dualism between reason and nature, since McDowell insists that the space of reasons is a (second) natural phenomenon, but it is a descendent of that dualism.

McDowell casts the distinction as one between two different “modes of intelligibility”. There is interpretation by appeal to reasons, on the one hand, and explanation by appeal to natural law, on the other. This view is reminiscent of the two-standpoints reading of Kant that I introduced at the end of the last chapter. Recall that Kant writes in the *Groundwork*:

…he [i.e. a human being] has two points of view from which he can regard himself, and recognize laws of the exercise of his faculties, and consequently of all his actions: first, so far as he belongs to the world of sense, he finds himself subject to laws of nature; secondly, as belonging to the intelligible world, under laws which being independent of nature have their foundation not in experience but in reason alone (Kant, *Gr*: 113).

In McDowellian terminology, we see ourselves from two different explanatory standpoints or modes of intelligibility. We can view ourselves as parts of the natural order, subject to causal explanation. In contrast, we can also see ourselves in terms of
rational intelligibility, as rational agents operating within the space of reasons. The adjustment McDowell makes to the Kantian picture is that he insists that, if we adopt a suitably broad conception of nature, we are not forced to hold that the space of reasons is a non-natural or supernatural realm.

It is important to recognize here that the different modes of intelligibility do not apply to separate events or objects. In this sense, it is not as if we can somehow divide or categorize our actions as falling into either the mode of law-like intelligibility or rational intelligibility. The point, instead, is that we can view the same action in both modes. We can think of the same action, and coherently explain the same action, with reference to either type of explanation—through an appeal to scientific principles or through an appeal to rational explanation. Take human movement, for example. Let us say that I raise my hand in class. According to the dualistic framework proposed by McDowell, this bodily movement is subject to both forms of explanation. We may first of all appeal to natural intelligibility to explain the movement, and thus see the raising of my hand as an event in a greater causal chain. This sort of explanation for my hand’s rising makes no reference to me, as an intelligible agent capable of thinking, judging, deciding, and doing. The event, my hand rising, is explained by reference to whatever antecedent physical events eventually resulted in this motion. This same motion, however, may also be subject to rational explanation. This explanation makes essential reference to my reasons for raising my hand. Thus the rational explanation has to do with the fact that I have a query, and I judge that if I raise my hand the query will be answered. As such I decide, on these bases, to perform the action. Unlike the causal explanation, therefore, the
rational explanation refers to the reasons I have for raising my hand to explain my hand’s rising.

We now confront the critical question of how to understand the relationship between the two modes of intelligibility. McDowell holds that both forms of explanation are unproblematically compatible with each other. Explanation in terms of reasons cannot be reduced to explanation cast in terms of law: there is no possibility of a scientific-naturalistic reduction of the space of reasons. McDowell’s reasons asserting the autonomy of the two modes of intelligibility are basically \textit{a priori}. He holds that the psychological vocabulary in which we cast rational explanations cannot simply be reduced to physical vocabulary because the latter is unable to preserve the inferential structure that surrounds beliefs and desires. If we ascribe an intentional state to an individual, with this ascription we can infer a variety of other states. For example, let us ascribe to Jones the following belief state: “Jones believes that his car is running well”. With this, we can infer various other things: that Jones believes \textit{something} is running well, that Jones believes a car is a vehicle, that Jones believes he owns a car, and so on forth and so on. The reductionist is in a bind here, for it seems that for each ascribable belief we will have to ascribe a different mental state, thus multiplying the number of mental states ascribed without a physical basis. The connection between the numerous psychological states, the ascription of which follow from the ascription of any one does not appear to be a physical connection. It is a matter of rational relations of inference.

A determined naturalist, however, may take a very different view of the matter. She will argue that the issue of the relation of the two modes of intelligibility is not to be
settled *a priori*. It is ultimately an empirical question whether explanations cast in terms of reasons can be reduced to something scientifically more respectable. After all, the naturalist may say, human beings evolved rational modes of intelligibility in utter ignorance of the real physical basis of their mental lives, so it is hardly surprising that there seem to be essential incompatibilities between rational and law-like explanation. But we ought not to rule out *a priori* the possibility that, as our scientific expertise grows, we will find compelling scientific explanations of the behavior presently explained by appeal to rational relations. In other words, the reductionist will say that the more exhaustive the realm of law explanation becomes, the less the space of reasons will appear to be autonomous.

The naturalist may take the view that explanations cast in terms of scientific law will be able to capture everything worth capturing about rational explanation. Or she might take a more radical view which has come to be known as “eliminative materialism”. This view is a 20th century theory suggested by elements in the work of such as Wilfrid Sellars and W.V.O Quine, and taken up and developed by Paul and Patricia Churchland.10 The theory makes the radical claim that our ordinary and commonsense understanding of the mind, what eliminative materialists classify as “folk psychology”, is extensively mistaken. Our understanding of our psychological states and processes—as intentional states, beliefs, and desires—is simply wrong. These states and processes will have no place, at any level of analysis, in a sophisticated and accurate account of the mind. It does not matter, therefore, if the scientific-naturalist fails to
preserve the inferential structure of intentional beliefs or make sense of the relationship between mental states because, according to the eliminative materialist, this theory of mind will eventually be replaced by a neurological story that no longer refers to these types of states and processes. The original McDowellian response to the reductionist threat, therefore, is vacuous.

We need, then, to descend from abstract speculation about whether explanations cast in the terminology of the realm of law might eventually subsume those cast in terms of the space of reasons and find some more concrete way to consider the ways in which our conception of ourselves as free rational agents might be threatened by the deterministic explanations of natural science. Just such a discussion can be found in Peter Strawson’s famous essay, “Freedom and Resentment”, which is the subject of my next chapter.

10 See Ramsay, 2003 for a fuller account of the historical development of eliminative materialism.
Chapter 3

(i) Framing Strawson’s argument: optimists and pessimists on moral responsibility and determinism

Strawson’s goal, in his essay “Freedom and Resentment”, is to show that the long-standing debate between compatibilists and non-compatibilists regarding moral responsibility and determinism (whom he calls the “optimists” and “pessimists” respectively) is misguided. Section (i) of this Chapter outlines this long-standing dispute. In sections (ii), (iii), and (iv), I turn my attention to the body of Strawson’s essay and his central claim—that a reductivist account, (specifically, physical determinism) is irrelevant to our concepts of morality and moral responsibility, since these concepts are grounded and justified through what he terms participant reactive attitudes (Strawson, 1997: 119). The ultimate goal of this Chapter, in light of the criticism of McDowell developed in the previous chapter, is to show (in section (v)) that with Strawson’s help, McDowell is able to evade the charge laid down by the hard-nosed naturalist. The space of reasons can be rescued!

As mentioned above, Section I of Strawson’s essay introduces the topics of moral responsibility and determinism through a dispute between the optimists and the pessimists. The pessimists are those who assert that the truth of determinism would undermine the concepts of moral obligation and responsibility, blame and punishment. In opposition, the optimists assert that these same concepts would not lose their value or function if determinism were true. To this end, they invoke the value that these concepts
have in regulating social behavior, which is surely compatible with the truth of
determinism. Strawson notes that the pessimists would respond that: “…just punishment
and moral condemnation imply moral guilt and guilt implies moral responsibility and
moral responsibility implies freedom and freedom implies the falsity of determinism”
(Strawson, 1997: 120).

According to Strawson, the optimists would reply in turn by conceding that
punishment and condemnation do require freedom, but that we do possess freedom in the
requisite sense. Here they speak of negative freedom—“the absence of certain conditions
the presence of which would make condemnation or punishment inappropriate”—like
compulsion by another, innate incapacity, or insanity (Strawson, 1997: 120-1). The
pessimists would counter that to truly justify condemnation or punishment we must look
beyond negative freedom to positive freedom. This latter type of freedom requires not
only that an agent be free from the sorts of conditions just named, but also that they act in
accordance with their will’s desire, what Strawson characterizes as: “a genuinely free
identification of the will with the act” (Strawson, 1997: 121). But the optimist, Strawson
contends, would simply respond that this sort of freedom is included in their account, for
people surely do things, really intend to do what they do, and know just what they’re
doing in doing it (Strawson, 1997: 121). Finally, the pessimists would complain that the
truth of determinism cannot be reconciled with the idea of genuine free acts of will.

According to Strawson, the mock dialogue between the pessimists and optimists
results in a stalemate. The optimists are unable to convince the pessimists that their kind
of freedom is enough, and thus it looks as though the only condition that they have to
justify condemnation and punishment is the social desirability and efficacy of these institutions. On the other hand, the pessimist seems to be committed to the idea of the will as literally un-determined by anything other than itself, as standing aloof from nature, and this seems to inherit all the problems we rehearsed in our discussion of Kant. Strawson ends Section II by posing a question: “…might we not induce the pessimist to give up saying this by giving the optimist something more to say?” (Strawson, 1997: 122) The next sections, in both Strawson and this Chapter, are devoted to this “something more”.

(ii) Participant reactive attitudes

According to Strawson, the dispute just outlined reflects a detached and impersonal view of moral responsibility, punishment, and condemnation. In Section III of his essay Strawson proposes changing directions for the moment, “…to speak…of the non-detached attitudes and reactions of people directly involved in transactions with each other; of the attitudes and reactions of offended parties and beneficiaries: of such things as gratitude, resentment, forgiveness, love, and hurt feelings” (Strawson, 1997: 122). Unlike what Strawson refers to as our “cool, contemporary style” of philosophy, he proposes to look at our participant reactive attitudes—those attitudes that arise out of “…how much we actually mind, how much it matters to us, whether the actions of other people …reflect attitudes towards us of good will, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other” (Strawson, 1997: 123). Here the idea is that a person’s intentions towards another affect how the other reacts to the
resulting action—these are the participant reactive attitudes. To give one of Strawson’s examples, I may feel nothing toward another whose actions functioned to benefit me without his knowledge. However, if the person’s actions were performed intentionally to benefit me, out of his good will towards me, I shall feel gratitude towards him. In this case then, gratitude is the participant reactive attitude.

In terms of these attitudes Strawson wants to begin his inquiry by looking at two things: (i) those situations in which the attitudes are inappropriate, and (ii) what it is like not to experience the attitudes. Strawson suggests that we start by looking at the attitude of resentment, which arises from: “…situations in which one person is offended or injured by the action of another” (Strawson, 1997: 124). More specifically, he wants to consider special cases in which the attitude ought to be nullified. Strawson contends that the first type of these sorts of situations can be characterized by expressions such as “He didn’t mean to”, “He hadn’t realized”, “He didn’t know”, and “He couldn’t help it” (Strawson, 1997: 125). Though Strawson concedes that these expressions could apply to various different situations and scenarios, he argues that they share a commonality, namely: “None of them invites us to suspend towards the agent, either at the time of his action or in general, our ordinary reactive attitudes” (Strawson, 1997: 125). These sorts of situations are thus ones in which we do not cease to see the agent as a fully responsible human being, though in these particular instances we would not deem him responsible for whatever injuries or offenses occurred. In other words, because of the considerations noted above, these sorts of situations function to inhibit the injured party’s resentment
towards the other agent, but this inhibition does not involve a general suspension of all participant reactive attitudes towards to offending or injurious agent.

Strawson proceeds to consider the next type of situation in which the attitude of resentment ought to be nullified. He categorizes these situations into two separate subgroups. The first subgroup is characterized by expressions such as: “He has been under very great strain recently”, “He wasn’t himself”, “He was acting under post-hypnotic suggestion” (Strawson, 1997: 126). The second subgroup, which Strawson affirms as the more important group for the present discussion, is characterized by expressions such as: “He’s only a child”, “He’s a hopeless schizophrenic”, and “His mind has been systematically perverted” (Strawson, 1997: 126). These last expressions reflect situations in which the circumstances of the situation are normal but the acting agent is psychologically abnormal or morally immature (so unlike in the first sorts of cases, these situations reflect those in which the agent is generally subject to either condition—not momentarily subject to whatever stressor). According to Strawson, these situations are so important because they reflect cases wherein we typically suspend or nullify all ordinary participant reactive attitudes and adopt what he terms the objective attitude. Strawson writes:

To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide

---

11 Although Strawson is grouping the psychologically abnormal agent and the morally immature agent together here, it should be noted that the latter (a child) is different from former insofar as he or she will normally grow out of this general condition (that of being morally immature) as he or she ages and develops. Though the psychologically abnormal agent may return to normalcy at some point (a schizophrenic can be certainly be treated, to a certain extent), he or she remains different from the child insofar as it is less likely for him or her to “grow out of” the condition, as a child typically should.
range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained...The objective attitude may be emotionally toned in many ways, but not in all ways: it may include repulsion or fear, it may include pity or even love, though not all kinds of love. But it cannot include the range of reactive feelings and attitudes which belong to involvement of or participation with others in inter-personal human relationships; it cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally (Strawson, 1997: 126-7).

According to Strawson, adopting the objective attitude is what we typically do with psychologically abnormal and morally immature agents, but it is also something we can do with normal, mature adults. Unlike with the psychologically abnormal or the morally immature, however, this is not the typical form of our interactions with other normal mature adults. It is something we may occasionally and temporarily adopt “as a refuge...from the strains of involvement; or as an aid to policy; or simply out of intellectual curiosity” (Strawson, 1997: 127). So although we may employ the objective attitude at certain times, this is not the habitual form of our interactions with other like agents. Instead, we are naturally predisposed towards the participant reactive attitudes in our day-to-day interpersonal relationships.

Having drawn this distinction, between participant reactive attitudes and the objective attitude, Strawson return to our original query having to do with the effect of determinism on moral responsibility. Recall for the moment that unlike the optimists and pessimists, who speak of moral responsibility in terms of the formal institutions of condemnation and punishment, Strawson is here talking about the nature of our
interpersonal relationships—those attitudes that arise out of one’s perception of another’s will towards him. With this in mind, Strawson poses the following question:

What effect would, or should, the acceptance of the truth of a general thesis of determinism have upon these reactive attitudes? More specifically, would, or should, the acceptance of the truth of the thesis lead to the decay or the repudiation of all such attitudes? Would, or should, it mean the end of gratitude, resentment, and forgiveness; of all reciprocated adult loves; of all the essentially personal antagonisms? (Strawson, 1997: 127).

Before addressing this question Strawson states that we must first make headway regarding what the thesis of determinism actually means. He claims there is one thing we do know about the meaning of it, namely: “…that if there is a coherent thesis of determinism then there must be a sense of determined such that, if that thesis is true, then all behavior whatever is determined in that sense” (Strawson, 1997: 128). With this in mind, Strawson contends that we can consider what possibilities lie open to us without knowing exactly what the thesis of determinism entails, since we have before us a sketch of the various participant reactive attitudes, and those situations in which they ought to be inhibited.

Strawson begins with the first group of considerations discussed earlier, those in which the attitude of resentment ought to be inhibited towards a particular agent in specific circumstances but without inhibiting reactive attitudes in general towards the agent in question. He asserts that these sorts of situations are not suitable for answering our present question since we want to know about just this: the general inhibition of reactive attitudes. Next Strawson turns to the second sorts of situations, those in which the objective attitude is assumed towards psychologically abnormal or morally immature
agents. These situations as well, Strawson contends, are not relevant to our question since although they involve the suspension of reactive attitudes generally, this general suspension applies only to limited members of any given society. Stated by Strawson, “…it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition” (Strawson, 1997: 128).

Strawson therefore turns his attention to the last sorts of situations discussed earlier, those in which we can sometimes look on normal, mature agents through the objective attitude, whether it be for avoiding the strains of involvement, making social policy, or simply for curiosity’s sake. This, he claims, is the condition worth considering in terms of the effect that the truth of the determinist thesis would have on it. Our question becomes, therefore: “Could, or should the acceptance of the determinist thesis lead us always to look on everyone exclusively in this way?” (Strawson, 1997: 129).

Strawson answers this question in two parts. The first part of his answer appeals to intuition. He states that although it is not contradictory to suppose this might happen—to suppose that that truth of determinism would lead to a state of affairs in which everyone comported themselves in accordance with the objective attitude—that this seems highly unlikely. To quote Strawson:

The human commitment to participation in ordinary interpersonal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them; and being involved in inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question (Strawson, 1997: 129).
This analysis seems decidedly accurate. Though it is true that we cannot prove this claim with certainty (so show for sure that the truth of determinism would not lead to a universal adoption of the objective attitude), this proposal looks implausible. The emotions and feelings that make up our participant reactive attitudes and form the foundations for our inter-personal relationships are a natural part of our humanness—not things that normal mature adults may or may not acquire, or dispose of at will. As Strawson points out, adopting the objective attitude is something that we can sometimes do towards other like adults, but doing this is not something that can be sustained or prolonged easily, if at all. Adopting this attitude goes against a fundamental feature of our humanity. In this sense it seems inconceivable that a theoretical proof of determinism would necessarily lead the adoption of such an unnatural mode of behavior.

The further point that Strawson makes here is that our adoption of the objective attitude, whether it is in cases of the psychologically abnormal agent, the morally immature agent, or the short-term adoption of the attitude towards a normal mature adult, is never the result of a theoretical conviction in the thesis of determinism. Instead, this attitude is adopted because we choose to suspend our typical participant reactive attitudes (for various reasons differing widely between unique situations). The point here is that it not only seems unlikely that we would widely adopt the objective attitude given a conviction in determinism, but that our present usage of the objective attitude does not result from this conviction anyway.

Strawson then considers whether or not it would be rational to adopt the objective attitude given a general conviction in determinism. In response he argues that those who
pose this question have failed to understand the importance of the preceding discussion, the idea that the participant reactive attitudes are natural, fundamental features of our humanity. Stated by Strawson: “This commitment [to the participant reactive attitudes] is part of the general framework of human life, not something that can come up for review as particular cases can come up for review within this general framework” (Strawson, 1997: 130). Moreover, Strawson claims, even if we could imagine a world in which we had a genuine choice as to whether or not to adopt the objective attitude, the rationality of this choice would not rest on whether or not determinism was true. Instead, as stated by Strawson: “…we could choose rationally only in the light of an assessment of the gains and losses to human life, its enrichment or impoverishment; and the truth of falsity of a general thesis of determinism would not bear on the rationality of this choice” (Strawson, 1997: 130).

(iii) The vicarious analogues

In Section V Strawson turns to a discussion of what he terms the ‘vicarious analogues’ to the various sorts of participant reactive attitudes we have been discussing thus far. To explain, the vicarious analogues refer to: “…reactions to the qualities of others’ wills, not towards ourselves, but towards others” (Strawson, 1997: 131). Strawson notes that the vicarious analogue of someone who feels resentment is the person who feels indignant or disapproving. The sorts of situations in which this particular analogue might arise are ones in which an observing agent experiences another being treated in an offensive or injurious way. Though they are not the personal recipient of the injury or offense, they
feel “resentment on behalf of another” (Strawson, 1997: 131). Unlike the participant reactive attitudes then, where one is concerned with another’s willing in respect of himself, the vicarious analogues reflect our general concern with another’s willing in respect of all others. Because of the generality that the vicarious analogues express, Strawson claims that these attitudes take on an important moral dimension. To clarify this distinction with an example, when I feel the participant reactive attitude of resentment, this expresses merely my perceived and personal offense or injury at the hands of another. When I feel its vicarious analogue, however (indignation or disapproval on behalf of another), this attitude reflects a general view about how it is appropriate for one person to treat another—it goes beyond the scope of the personal. In this sense we have entered the moral sphere.

As in the last discussion surrounding participant reactive attitudes, Strawson proceeds by inquiring into what sorts of considerations or special circumstances ought to inhibit the presence of the vicarious analogues. As we will see, the cases in which the vicarious analogues ought to be inhibited parallel those discussed earlier in reference to the participant reactive attitudes. So returning to our first sort of case, we now have states of affairs in which indignation or disapproval (instead of resentment) ought to be inhibited in situations characterized by expressions such “He hadn’t realized” or “He

---

12 At this point in his essay Strawson also mentions “self-reactive attitudes”, which complete the triad of attitudes associated with interpersonal relationships (1. participant reactive attitudes, 2. the vicarious analogues, and 3. the self-reactive attitudes). This latter type refers to those demands we place on ourselves for others. Feelings associated with self-reactive attitudes are those of obligation, guilt, shame, and remorse. Though Strawson briefly touches on these self-reactive attitudes in his essay, I choose not to include this discussion in my text proper since he proceeds quite quickly in his argument with reference to the vicarious analogues only.
didn’t know”. Like with the participant reactive attitudes, these same situations do not call for a full suspension of all the vicarious analogues of the reactive attitudes, but call for the suspension of indignation and disapproval given the special circumstances.

Keeping in mind the newfound moral dimension of the vicarious analogues, Strawson further claims that we can now refer to these sorts of situations as those in which the acting agent is not morally responsible for the injury or offense, though he is still generally viewed as a morally responsible agent.

Moving onto the cases of the psychologically abnormal or morally immature agent, Strawson asserts that just as is the case with the participant reactive attitudes, we tend to suspend or inhibit all their vicarious analogues when dealing with these types of agents. Moreover, due to the moral dimension that the vicarious analogues bring, we can additionally say that these sorts of agents are generally not morally responsible for their actions, not members of the moral community. Lastly, we may also suspend the vicarious analogues to the participant reactive attitudes for short periods of time in our dealings with normal mature adults. Strawson contends that this may be easier to do with the vicarious analogues, though our motives for doing as much are diminished. To explain, as Strawson noted earlier with reference to the participant reactive attitudes, I may take the objective attitude towards another to avoid the strains of involvement in some situation. This same motivation is not applicable when we look at the vicarious analogues however, since by contrast the participant reactive attitudes, they do not presuppose personal involvement in the situation. More generally, the claim is that there
is less tension between the moral vicarious analogues and the objective attitude, as opposed to between the participant reactive attitudes and the objective attitude.

Paralleling his last discussion, Strawson turns his attention to the question of whether or not a general thesis of determinism would have any impact on the vicarious analogues, the moral counterparts to the participant reactive attitudes. He begins by stating that the suspension of the vicarious analogues in the sorts of cases described above is never a consequence of a theoretical conviction in determinism. Stated by Strawson: “For it is not a consequence of any general thesis of determinism which might be true that nobody knows what he’s doing or that everybody’s behavior is unintelligible in terms of conscious purposes or that everybody lives in a world of delusion or that nobody has moral sense” (Strawson, 1997: 135). In terms of the possible social impact that a general conviction in determinism might have on the vicarious analogues, Strawson argues that it is hard to imagine a world in which the participant reactive attitudes continued to play a role and their counterparts disappeared. The participant reactive attitudes and their vicarious analogues are deeply connected to one another, or as Strawson puts it: “…as general human capacities…they stand or lapse together” (Strawson, 1997: 135). Insofar as the attitudes and their analogues are thoroughly connected, and insofar as Strawson has already argued that we cannot envisage a world without the various participant reactive attitudes given a general thesis of determinism, he concludes that the vicarious analogues would likewise not be impacted by such a theoretical conviction. Lastly, Strawson considers whether or not it would be rational to get rid of the vicarious analogues given a general conviction in determinism. As in the
case of the earlier discussion regarding the participant reactive attitudes, Strawson states that someone who asks this question has failed to grasp the importance of the discussion at hand. In his own words: “It is useless to ask whether it would not be rational for us to do what it is not in our nature to (be able to) do” (Strawson, 1997: 135).

Strawson concludes Section V by stating that thus far in his arguments he has merely provided a schema, marked by a crude distinction between the reactive attitudes and the objective attitude. He wants to briefly consider certain cases that oscillate between these attitudes in order to strengthen his point. The first case he considers is that of children. As I mentioned earlier, our treatment of children takes the form of the objective attitude, though with the knowledge that they will eventually grow into mature beings capable of the full range of participant reactive attitudes. What bearing, Strawson asks, would a thesis of determinism have on our treatment of children as such? According to Strawson, it would be “grotesque” to think of a child’s behavior, given a thesis of determinism, as transitioning from determined to undetermined. He next considers the relationship between a psychoanalyst and his patient. In these sorts of cases, the psychoanalyst adopts the objective attitude towards his patient, though the eventual purpose of this practice is to make the suspension of the ordinary reactive attitudes unnecessary. As stated by Strawson: “…the goal of this sort of enterprise means bringing it about that the agent’s behavior shall be intelligible in terms of conscious purposes rather than in terms only of unconscious purposes” (Strawson, 1997: 137). Once this goal is achieved, the objective attitude is dropped. The point therefore, with both the raising of a child and the treatment of a psychiatric patient, is that the idea of
being determined has nothing to do with our usage of both the reactive attitude and the objective attitude—including our transition from one to the other. Determinism is irrelevant to our adoption of both the reactive attitudes and the objective attitude, but also irrelevant to cases in which we oscillate between attitudes.

(iv) Return to the optimists vs. pessimists: Strawson’s ‘something more’

We can now turn to the last section in Strawson’s essay, and return to the original debate between the optimists and pessimists. Up for debate, to re-iterate, is the question of whether or not the truth of the thesis of determinism should have an effect on our concepts of morality and moral responsibility. According to Strawson, we can now see more clearly into the nature of the debate, and recognize that both the pessimists and the optimists are suffering from a similar misunderstanding. In Strawson’s words: “…in different ways they seek to over-intellectualize the facts” (Strawson, 1997: 140). The optimists do this through what he terms a “one-eyed Utilitarianism”, wherein the practices of holding persons morally responsible rest on the social efficacy of such practices. The pessimist recoils from this picture, in what Strawson characterizes as an intuitive emotional shock. The shock results from the fact that the optimist’s schema reflects merely the objective attitude. It excludes all the reactive attitudes from the concept of moral responsibility, turning morality into something merely for policy, treatment, and social control.

According to Strawson, the pessimists are quite right to feel emotional shock at this picture, but they too over-intellectualize the facts in response to the optimist
proposal. Though the pessimist is right not to lose sight of the reactive attitudes, he is unable to accept the fact that it is just these attitudes which fill the gap in the optimist account. In Strawson’s words: “Instead, he thinks the gap can only be filled if some general metaphysical proposition is repeatedly verified, verified in all cases where it is appropriate to attribute moral responsibility” (Strawson, 1997: 140). Though he only touches on this point briefly, the point is that the pessimist over-intellectualizes here by filling the gap with some sort of unnecessary metaphysical theory or proposition.\textsuperscript{13}

Though both theories are wrong to a large extent, insofar as they both seek to over-intellectualize the facts, Strawson concedes that the optimist is more right in terms of our original query, having to do with the effect of the truth of determinism on morality. It is right to say that determinism does not undermine the concepts of morality and moral responsibility because of the social efficacies that the concepts bring. More importantly, however, is the idea that the thesis of determinism does not undermine the concepts of morality and moral responsibility because we are naturally equipped with participant reactive attitudes and their vicarious moral analogues, attitudes that express our fundamental commitment to a moral life. As we have seen through Strawson’s earlier discussion, the participant reactive attitudes as well as their suspension cannot be shaken by determinism. It does not look as though the truth of the thesis of determinism would have any effect on these deeply ingrained sentiments. Moreover, the adoption of the objective attitude is never a consequence of the acceptance of this thesis. In this sense,\textsuperscript{13}

\begin{flushright}
\textsuperscript{13} In the text Strawson briefly refers to two such propositions: “contracausal freedom” and “intuitions of fittingness” (Strawson, 1997: 140). For my present purposes an examination of such metaphysical propositions is unnecessary.
\end{flushright}
the optimist is on the right track, but the pessimist is also right that our status as rational beings somehow goes missing if we adopt the objective standpoint the optimist favors, and this we can see by reflecting on the reactive attitudes and their moral analogues. As Strawson puts it: “Our practices do not merely exploit our nature, they express them” (Strawson, 1997: 142). The concepts of morality and moral responsibility, therefore, apply to us regardless of a general thesis on determinism—not only because of their social role, but because they reflect how we really feel, how we really are.

(v) A return to McDowell

We can now return to McDowell, in order to see specifically how Strawson’s proposal in “Freedom and Resentment” fills out McDowell’s account of the space of reasons presented in Mind and World. As I argued at the end of Chapter Two, it is not clear that McDowell really has a non-question-begging argument against the hard-nosed naturalist—he simply insists that the kind of intelligibility conferred by space of reasons explanations cannot be captured in terms of the realm of law. What Strawson offers McDowell is an argument to the effect that there is a limit to the extent to which the objective viewpoint implicit in the realm of law can do to displace the normative explanation of the space of reasons. We cannot get rid of the space of reasons explanation, Strawson shows, even if we could somehow become convinced that exhaustive scientific explanations of human behavior were available. Focusing specifically on the concepts of freedom and moral responsibility, Strawson argues that both sides of the traditional debate surrounding the compatibility of determinism with
these concepts are misguided. Instead Strawson focuses on the sentiments and attitudes
that make up the basis of our everyday, interpersonal relationships. His claim is that our
predisposition towards these participant reactive attitudes, as well as their vicarious moral
analogues, demonstrates our commitment to a moral life. They represent the fact that we
operate within the space of reasons regardless of a genuine commitment to a theoretical
truth of determinism. In this sense, although explanations cast from the objective
standpoint might in certain particular cases displace explanations in terms of the agent’s
reasons, no general account cast from the objective viewpoint could ever undermine our
conception of ourselves as inhabitants of the space of reasons.

In sum then, Strawson reinforces the McDowellian idea that the correct strategy
in terms of reconciling freedom and nature is a compatibilist one. Unlike Kant himself,
who posits noumenal freedom in the service of incompatibilism, both McDowell and
Strawson think that the mode of intelligibility of the space of reasons in nature and that of
the deterministic framework of the realm of law are essentially compatible. McDowell,
however, is focused on quite an abstract issue—namely, whether a baldly-naturalistic
philosophical theory of meaning on normativity is possible. In contrast, Strawson really
brings things down to earth. Unlike McDowell, he is not interested in the plausibility of
broad explanatory strategies, but in how we come to understand and relate to one another
in everyday discourse. By focusing on this latter issue, Strawson shows that although
McDowell is quite right to locate the space of reasons in nature, we need not defend this
space by reference to the incoherence of a reductionist theory. For what Strawson shows
is that even if such a theory were proved true, this could not impact, at least on a
widespread scale, our commitment and necessary existence as beings in the space of reasons.

The position at which we have arrived can be seen as a blending of Kantian and Humean insights. Like Kant, McDowell and Strawson relate freedom to our status as accountable rational agents. However, the reactive attitudes on which Strawson focuses are precisely not the sort of states Kant thought of as expressive of this status, for they are natural reactions elicited by circumstances and are by no means in the domain of pure spontaneity. They are states of the kind Hume placed at the foundation of morality, not Kant. Nevertheless, the reactive attitudes are within the space of reasons as McDowell understands it because they are open to rational reflection and normative assessment. We can ask whether it is appropriate for us to feel resentment, argue that someone ought to feel gratitude, and so on. Although these states are expressions of our nature, they should not therefore be thought antithetical to reason: on the contrary, McDowell's whole point is that we need a conception of reason as within nature and that we must therefore broaden our conception of the natural beyond anything Kant or Hume would have acknowledged.

(vi) Conclusion

This thesis has argued the following. In his effort to provide a rational justification for causal judgments against Humean skepticism, Kant finds himself committed to the view that all events in the world we experience are subject to causal law. Since Kant’s moral philosophy commits him to the view that the reality of genuine freedom is a precondition
of moral responsibility, his only option is to locate freedom in the shadowy noumenal world of things in themselves. This is a profoundly unsatisfactory solution. A more plausible position, still Kantian in spirit, is to be found in the philosophy of John McDowell. McDowell agrees with Kant that our freedom consists in our status as rational beings. He argues that philosophers have been compelled to represent our rationality as somehow outside nature because they have inherited an unduly restrictive conception of the natural from the scientific worldview of the 17th century. McDowell argues that if we extend our conception of nature to include the second-natural capacities that human beings acquire in the course of maturation and education, then we can acknowledge that our status as free, rational agents is wholly natural. McDowell maintains, however, that our standing as rational agents is not intelligible in natural-scientific terms. In a way reminiscent of Kant, he argues that there are two distinct modes of intelligibility: rational explanation and explanation by appeal to scientific law. Our status as free agents is disclosed by the former mode of explanation, our status as merely-natural beings by the latter. McDowell holds that there are *a priori* reasons to believe that the two modes of intelligibility are autonomous and that a reductive or eliminativist explanation of the “space of reasons” can never be given in scientific terms. A hard-nosed naturalist, however, will not be persuaded by the kind of general *a priori* considerations McDowell invokes. However, another contemporary philosopher inspired by Kant, Sir Peter Strawson, provides an ingenious account of why there is no prospect that our understanding of ourselves as rational agents will somehow be subsumed by explanations from the objective standpoint of deterministic natural science.
By focusing on “reactive attitudes” such as resentment and gratitude, Strawson examines in detail the circumstances in which we suspend those attitudes in the face of considerations about the causation of behavior, and concludes that the truth of determinism (and the availability of extensive natural-scientific explanations) is irrelevant to the cogency of the familiar forms of explanation that are central to our conception of ourselves and our relations to others. Moreover, our commitment to the reactive attitudes is symptomatic of our acting within the space of reasons, for our reactive attitudes are open to rational deliberation. Our usage of the reactive attitudes is open to the question of whether or not they ought to be used, or whether or not they are appropriate in particular circumstances. They are thus constitutive of McDowell’s naturalized space of reasons. The result is a persuasive form of compatibilism that reconciles freedom and determinism, and is Kantian in spirit without the problematic elements of Kant’s own position with its reliance on the obscure realm of the noumenal.
Bibliography

Kant Texts


Other Texts


