

SUBJECT, NORMATIVITY, WORLD

by

Stefan Bird-Pollan

Dissertation

Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Philosophy

May , 2008

Nashville, Tennessee

APPROVED:

Gregg Horowitz

John Lachs

David Wood

Jay Bernstein

TABLE OF CONTENTS

Page

INTRODUCTION	4
THE STRUCTURE OF THE DISSERTATION	8
Chapter	
I. AUTONOMY AND KANT’S (FAILED) DEDUCTION OF FREEDOM	13
PRACTICAL REASON, AUTONOMY AND MORAL INSIGHT.....	15
KANT’S DEDUCTION.....	18
THE ANALYCITY THESIS AND RECIPROCITY THESIS.....	19
THE RECIPROCITY THESIS.....	21
THE PREPARATORY ARGUMENT AND THE ARGUMENT AGAINST FATALISM.....	25
THE HIDDEN CIRCLE	28
THE TWO STANDPOINTS ARGUMENT.....	29
THE CATEGORICAL IMPERATIVE	31
PRAUSS: PURE PRACTICAL REASON AND MORALITY	39
PRACTICAL REASON AND AUTONOMY.....	42
KAULBACH: THE HISTORICAL PERSPECTIVE ON AUTONOMY.....	46
CONCLUSION.....	54
II. RAWLS AND CONSTRUCTIVISM	57
THE ORIGINAL POSITION AND THE CATEGORICAL IMPERATIVE	59
THE MORAL CHARACTER OF THE IDEAL AGENT.....	64
THE REASONABLE AND THE RATIONAL.....	66
CONSTRUCTIVISM.....	74
EXCURSUS: RAWLS AND KANT, PARALLEL ARGUMENTS	80
JUSTIFICATION AND THE REFLECTIVE EQUILIBRIUM.....	83
PROBLEMS WITH THE REFLECTIVE EQUILIBRIUM	89
PURE PRACTICAL REASON AND THE FIRST PERSON PERSPECTIVE	93
THE PROBLEM OF NORMATIVITY AND THE NECESSITY OF ITS JUSTIFICATION	97
III. HEGEL’S AND WILLIAMS’ CRITIQUES OF THE KANTIAN POSITION	103
THE CRITIQUE OF THE CATEGORICAL IMPERATIVE I: THE FORMALITY CHARGE.....	104
THE CRITIQUE OF THE CATEGORICAL IMPERATIVE II: THE FAILURE OF THE RECIPROCITY THESIS	108
THE PROBLEM OF INTERSUBJECTIVITY IN KANT AND RAWLS.....	112
A REORIENTATION OF THE NORMATIVITY QUESTION: JUSTIFICATION AND MOTIVATION	114
WILLIAMS’ CRITIQUE OF THE KANTIAN POSITION	115
REFORMULATING THE QUESTIONS AFTER HEGEL AND WILLIAMS	124
PROSPECTS FOR A REFORMULATION OF THE KANTIAN PROJECT.....	125

IV. KORSGAARD ON SELF-CONSTITUTION.....	128
THE QUESTION OF MORAL CHARACTER	129
PRACTICAL REASON AND REFLECTION.....	134
ACTING ON PRINCIPLES	138
ACTING ON THE PRINCIPLE OF PURE PRACTICAL REASON	141
THE NORMATIVE ARGUMENT	146
SELF-CONSTITUTION, NORMATIVITY AND TELEOLOGY.....	148
MOTIVATION AND GOOD OR BAD ACTION	151
CONCLUSION.....	155
V. INTERSUBJECTIVITY IN KORSGAARD	159
HEGEL’S ACCOUNT OF ACTION	160
THE SUBJECTIVISM CHARGE: CONSCIENCE AND GAUGUIN’S GROUND PROJECT	163
THE INTERSUBJECTIVITY OF REASON IN KORSGAARD: THE WITTGENSTEINIAN ARGUMENT	168
THE INTERSUBJECTIVITY OF REASON IN KORSGAARD: TAKING OTHERS’ REASONS AS ONE’S OWN	171
DELIBERATION AS ACTION	175
THE DIVIDED SELF	179
CONSTITUTIVE STANDARDS OF ETHICAL ACTION.....	182
CONCLUSION: THE QUESTION OF JUSTICE	185
VI. HEGEL AND REFLECTION.....	188
THE PROBLEM OF JUSTICE	191
HEGEL’S APPROPRIATION OF THE TRANSCENDENTAL UNITY OF APPERCEPTION	196
REFLECTION	197
SELF-CONSCIOUSNESS AND THE APPERCEPTION THESIS	201
THE MASTER-SLAVE DIALECTIC	205
SELF-POSITING THOUGHT AND CONTINGENCY.....	207
LINKING HEGEL’S SPECULATIVE AND PRACTICAL PHILOSOPHY	210
THE TRANSCENDENTAL THESIS AND HEGEL’S FAILURE.....	213
CONCLUSION.....	215
VII. HEGEL, INTERSUBJECTIVITY AND HISTORY.....	217
THE MASTER-SLAVE DIALECTIC: RADICAL FREEDOM AND SLAVERY	218
THE MASTER-SLAVE DIALECTIC: AGENCY AND HISTORY	222
PINKARD’S CATEGORIAL READING AND THE PERSPECTIVE OF THE SLAVE	234
FROM METAPHYSICS TO ONTOLOGY	238
SPIRIT’S PURPOSIVENESS	242
DIALECTIC, NEGATIVITY AND NECESSITY	243
FROM METAPHYSICAL TRUTH TO SOCIAL TRUTH.....	246
REVISABILITY: OUR HISTORY, OUR CONCEPTS	249
CONCLUSION.....	251
CONCLUSION	254
BIBLIOGRAPHY	263

INTRODUCTION

In the broadest terms, this dissertation is an investigation into what it would mean for practical reason to provide normative categories of intersubjective action in the service of universal freedom. Thematically, my strategy in pursuing this question will be first to examine the notion of practical reason itself and then to see whether practical reason as it is conceived of by Kant can be genuinely intersubjective. The concept of intersubjective action, understood as one agent's reason being a motivating reason for another agent, gives rise to two related questions: the first concerns whether it is possible at all and the second concerns how the possibility of intersubjectivity is related to justice.

Historically, my strategy will be to pursue these questions in the works of broadly Kantian writers, Kant, Rawls, Korsgaard and, lastly, Hegel. By working in this tradition I would like to argue that the internal pressures of the failure of Kant's original argument linking freedom and reason forces Kantian writers like Rawls and Korsgaard to move closer to the position advocated by Hegel and develop a more immanent understanding of the relation between world and subject. From a historical point of view, we can see this rapprochement also as the overcoming of the widely diagnosed metaphysical residue which plagued Kant's writing: the radical distinction between the free will of the subject and the deterministic world of physics.

I further believe that arguing for the continuity of the problematic of practical reason and immanent reflection on both the 'analytic' and 'continental' sides of the debate means that this once hard and fast distinction is becoming of less and less value. Convergence of the two sides can be seen in the fact that both Bernard Williams' and Hegel's critique of the Kantian position overlap in many aspects. But more important than that, I wish to emphasize that these two criticisms lead to a reshaping of the debate which in no way turns away from Kant's original commitments to freedom and intersubjectivity. I thus see Hegel's role in this dissertation not as adversarial but rather as completing a project which has recently been articulated more clearly by writers like Rawls and Korsgaard.

The aims of this dissertation can also be stated in slightly different terms, namely those of the now 200 year old debate between Kant and Hegel. As indicated, I see general agreement between Kant and Hegel in terms of the scope and ways of grounding a theory of practical reason. However, this should not blind us to the still significant differences between the Kantian and the Hegelian perspectives. These differences will be examined here from the agent-centered perspective. I will argue that the Hegelian perspective is that of the agent acting within history while the Kantian perspective is essentially ahistorical and as such fails to account for the position of the real existing individual's position.

This difference lies at the heart of the familiar claim that Kant's moral theory is empty. But to say only that is to miss the fact that Kant's notion of

freedom does not commit us to an ahistorical perspective. What I would like to draw out in this dissertation is the extent to which the perspective of individual agency might be retained within the Kantian account if this account is modified to emphasize the position of the individual striving for, but not necessarily achieving, universal justice. The decisive shift in the philosophical transition from a Kantian theory to a Hegelian one is accordingly to be found in rejecting the idea that a complete system of duties can be articulated from an impersonal perspective and in the embrace of the idea that the duties we do have are only ever the result of concrete and historically situated rational reflection.¹

This shift is decisive since it is only from the latter perspective that the contingencies of the world can be both accounted for and made less arbitrary with regard to justice. A merely ideal theory will fail to connect to concrete decisions and so be insufficient to orient an agent's thinking in the world. This point is powerfully made by both Hegel and Bernard Williams. Williams insists on the non-ideality of the present world and argues vehemently that in order to make sense of actions, we must take their origin in this non-ideal world into consideration. Universal prescriptions will not help us here.

Indeed, reason is able to abstract from its historical conditions by universalizing its aims but this does *not* mean that my current reflections are universal in the sense of being unchanging. In drawing up the parameters of my

¹ I do not want to claim that Kant himself held such a rigoristic theory. My discussion here does not concern itself with seeing whether the categorical imperative can be gotten to 'work' in some

reflections, I might be inattentive to certain features of my environment that future generations will consider of central importance— my carbon footprint for example. The agent centered perspective universalizes but does not consider its results universal because it does not believe that reason can completely outstrip contingent conditions. If it could, what good would its judgments on current situations be?

The key to a theory of justice, I will argue, involves the idea that one can and must continually revise ethical norms and political laws with an eye toward making them more and more universal though also with the recognition that no law or norm can last. Reflection is only ever for the short to medium term.

The issue which shadows and, I will argue, is deeply connected to, the question of the perspective of the agent and the justice of her actions is the issue of metaphysics. For the division between freedom and nature which Kant seeks to overcome is fundamentally an question of overcoming the idea that there is a physically determined order to the universe which humans cannot help but follow. To put it more pointedly, the question of justice is whether humans themselves can make sense of their suffering and thus whether they can do anything to alleviate it. The idea that there is or is not anything humans can do about their situation is essentially 'metaphysical' in the sense that it is a question that must go beyond physical nature. The idea of agency is thus in a strong sense itself

casuistic way. What is at stake here is the claim to universality of certain maxims deemed impermissible or enjoined.

metaphysical. And this means that the metaphysics of agency is the central topic of (meta-) ethics. Metaphysics thus asks after the sense of suffering and with that, it asks about the justice of some people suffering and not others.

With this claim I mean to draw attention to the fact that the problem of agency is constitutively tied to the problem of the meaning or sense of suffering for each individual. When I thus say that there is a change in perspective between the ahistorical Kantian position of rational reflection which assumes that all are capable of justice and the temporal Hegelian position that justice is something to be achieved both in terms of concrete social organization and philosophical reflection itself, I am pointing to a fundamental disagreement about the nature of metaphysics and how it can be overcome or discharged.

Thus the tendency in Kantian constructivism to sidestep the question of metaphysics by seeking to establish the possibility of agency misses precisely what gave rise to the question of metaphysics itself, namely actual suffering. Actual suffering, I will argue, can only be accounted for from the perspective of actual agents reflecting on their actual situation.

The Structure of the Dissertation

In the first chapter we will examine Kant's own attempt to show that humans are capable of autonomous moral action. This involves a deduction which takes the form of the argument that since we can think of ourselves as free, we must *be* free. Kant abandoned this attempt because he recognized that it relied

on metaphysically inadmissible notions of the ontological priority of the noumenal realm. In order to develop the problematic of self-grounding further, I will follow the analysis of Kant's deduction with two additional interpretations of Kant's practical philosophy, both of which suggest a way of overcoming Kant's difficulties. One will be the interpretation of Gerald Prauss who argues that Kant's failure to ground pure practical reason deductively stems from being forced to separate nature from the self. Prauss suggests that Kant might have found a way out of his difficulty if he had developed a more inclusive notion of autonomy. Heinz Kaulbach, by contrast, argues that history can play a decisive role in Kant's conception of practical philosophy. History can be seen as the gradual reorientation of the subject away from appetitive nature toward a commitment toward autonomy. Both of these approaches, I conclude, put pressure on Kant's strict division between the autonomous self and the causal world, requiring a more inclusive treatment of human desire in moral philosophy.

In chapter two, I consider Rawls' attempt to base a Kantian moral theory on the coherentist principle of the reflective equilibrium. Rawls' avowed intention to remain metaphysically neutral, however, presents a problem for the theory of intersubjectivity and therefore normativity since without the guarantee that the agents in Rawls' reflective equilibrium *do* share the same fundamental rational structure of acting on practical reason, there is no way of knowing what sort of a consensus one is building in negotiation. One might be building genuine consensus or one might just be talking at cross purposes while believing that genuine consensus is being built.

Chapter three examines two critiques of the Kantian enterprise of grounding the principle of pure practical reason. The purpose of this chapter is to give a basic outline of what the Hegelian critique of the notion of the categorical imperative as a test for the permissibility of action amounts to in Hegel and also to give the slightly different critique Bernard Williams makes of the possibility of having a 'rigoristic' morality system. In this chapter I argue that the critique of the Kantian enterprise voiced by Williams and Hegel comes down to a critique of the categorical imperative as a transcendental or reified rule for acting. That is, because the categorical imperative test is external to what it is testing, it cannot be a reliable guide to action. And, as such, to add Williams' other point, it burdens us with unreasonable demands which ask us to give up the fundamental commitments which make us who we are.

Against this, I present Korsgaard's theory of reflection in chapter four. Here I argue that because Korsgaard interprets the categorical imperative as the process of reflective endorsement of incentives back on themselves, there is no problem of a mismatch between the practical problem facing the agent and what the agent proposes as a solution. Showing this involves some discussion of Korsgaard's moral psychology. Ultimately I will argue that Korsgaard is an internalist about reason, though in a different way from Williams. I also show that the disagreement between Korsgaard and Williams about where fundamental values come from is a result of Williams' Humean internalism. But Korsgaard's argument that reflective endorsement provides the agent with reasons for action

does not yet mean that it provides others with reasons for action. This is what I call the egoism objection, already made against Rawls in chapter two.

In chapter five I follow up the on the problem of egoism by giving an account of Korsgaard's theory of the publicity of reason. Here she argues that since desires are in some way prior to their reflective endorsement, anyone's reasons can become candidates for endorsement. There is thus no strong distinction between my desires and other peoples' desires. I take this to have established the possibility of intersubjectivity or communal deliberation.

In chapter six, I turn to Hegel in order to give more depth to the notion of intersubjectivity by trying to see how a theory of intersubjectivity might be able to give us more concrete guidance for action than the idea of mere possibility of justice does. In this chapter, again using Williams, I argue that moving toward justice requires us always to already have certain standards of justice which are then displaced by the newly arrived at standards. I thus reinterpret Williams' challenge to the Kantian position as a deep concern about the unjust political organization that would force one to choose between fulfilling one's vocation, say, of becoming a great painter and staying with one's family. I thus argue that standards of justice are inherently historical.

In chapter seven, I provide an interpretation of Hegel's master-slave dialectic with a view to understanding Hegel's moral psychology. Here it is my aim to show that Hegel construes the development of society as a series of struggles in which the reasons of those previously unheard by the dominant class enter into the consciousness of those who hold power. I interpret Hegel as

insisting that justice is a matter of the historical revision of the standards for action.

I conclude the dissertation with some remarks about the connection between the process of philosophical reflection and the perspective of the agent within history.

CHAPTER I

AUTONOMY AND KANT'S (FAILED) DEDUCTION OF FREEDOM

Kant's moral philosophy, as we saw in the introduction, is based on finding a positive employment for pure reason. For Kant this means determining a way in which pure reason can guide our actions. This means showing not only that rational beings would conform to the law, but also that humans are rational beings. This double problem might be stated as the problem of bridging the gap between the weak autonomy thesis and the strong autonomy thesis, where the former thesis is simply the (analytic) notion that rational beings determine themselves according to reason and the latter thesis holds that humans *are* such rational beings. The transcendental deduction of morality in Kant's *Groundwork* chapter 3 is supposed to bridge this gap and thus show that human beings act under the law of pure practical reason.

Kant's problem, formulated in different terms, is the problem of whether reason can motivate us to action by itself or whether all action stems in some sense from 'natural' inclinations. To put it yet another way, the question is whether rational reflection can bring the results of the understanding's own synthesis of intuition under categories into accord with the regulative principles coming from reason itself. In this sense, the transcendental deduction of morality is really supposed to be an extension of the anti-metaphysical project carried out

in the *Critique of Pure Reason*, since the deduction of morality (just like the transcendental deduction of the categories) insists that morality (and knowledge of the external world) is independent of external givens, like god or society. Reason is here shown to be self-grounding. The problem is that the analogy between the deduction of the categories of the understanding in terms of the apperception thesis breaks down when it is applied to the realm of pure practical reason. The deduction of a principle of pure reason in the noumenal realm cannot then be translated back into a principle which has normative force in the matters of the sensible realm.

The deduction of the moral law was a failure as Kant himself came to realize in the *Critique of Practical Reason*. There Kant recognized that he had failed to show that reason is self-grounding in the way he wanted to. He failed to show that thinking of ourselves as moral is a sufficient (though necessary) condition for being moral. The problem, I will argue, is that in setting up the deduction, Kant's distinction between the sensible and the noumenal world comes back to haunt him. Once this division is invoked, there is simply no way of overcoming it. Showing this to be so will set the ground of a reworking of the notion of moral reflection in subsequent authors, all of whom seek to avoid Kant's noumenal-phenomenal distinction.

After a discussion of the failed deduction, I will turn to two influential interpretations of Kant's theory of morality, which point in directions not taken by Kant, but ones that will find their analogue in my discussion of Hegel. The first is Gerald Prauss' interpretation, which lays the foundation for an understanding of

Kant's moral theory as a post-metaphysical theory centering on the concept of autonomy rather than morality, much as Hegel's speculative theory does. Prauss claims that by running together morality and autonomy Kant reduces nature to a necessary evil while upholding an extra-worldly notion of morality. Second, I will examine Friedrich Kaulbach's attempt to establish a Kantian notion of autonomy by providing a historical structure to ethical reasoning in which autonomy is the site of a struggle between our inclinations and reason, and in which reason ultimately can win out and reorient the self toward reason. The claim, then, is that Kant's theory of action, if historically mediated, could provide a way of overcoming the problematic dichotomy of the noumenal and the sensible.

Practical Reason, Autonomy and Moral Insight

Before getting to the deduction of freedom and morality in *Groundwork* III it will be helpful to make some general remarks about what we are looking for in such a deduction. Let us recall again that the deduction of morality has two parts. The first shows that rational beings are moral (thus obeying the categorical imperative), and the second shows that humans are such rational beings. The first step thus has to do with autonomy but not yet with morality. It is only in the second step that moral is 'established' with recourse to an argument according to which humans have noumenal natures and are thus rational.

Another important element is the practical nature of reason. For Kant reason is practical when there are sufficient reasons for the will to realize a particular purpose. Reason, in other words, is practical when it is employed as

the answer to the question 'what should I do'? Practical reason is pure when it causes or motivates an action without reference to (natural) inclination. Acting from pure practical reason is thus acting autonomously. If practical reason is to be pure, it must satisfy the following two conditions of motivation and normativity: (1) reason must contain principles which explain what the will wants, since the will is only autonomous if it recognizes the right thing to do without reference to the world outside it; and (2) the will must also be binding. The self, and not nature, must be able to cause action.² Practical reason is thus the general capacity to realize those ends which the self sets for itself. In realizing its ends, the self must thus constantly negotiate with its natural part, its set of inclinations and the influences around it. But while action is always the result of some nexus of inclinations and deliberation, it is axiomatic for Kant's account of morality that only those actions are moral which are motivated without consideration for inclinations or external circumstances.

Understanding this type of autonomous self-relation is fundamental to understand the notion of practical reasons since it is not obvious that it can even be shown that the self has such a capacity for self-direction. German Idealism is based on the tenet that the self and its consciousness of itself are irreducible terms. This means that theoretical knowledge is knowledge of the self's knowledge. I know that I know. To show that practical reason is real, however, it must then be shown that this self-relation, this relation of the subject to itself can

² See Dieter Henrich. "Ethik der Autonomie". *Selbstverhältnisse*. Stuttgart: Reclam, 1982. pp. 6-56. 12-14.

also be understood in a practical way. This would mean that the practical self has to will itself in the same way it knows itself. But this would fundamentally involve the idea that theoretical knowledge is already practical, that knowing is already acting.³ As we shall see, the precise nature of the parallel between practical reason and theoretical reason will be an interpretative question. This is the role of the categorical imperative which seeks to unite knowledge with normative reasons for acting, but the difficulties in getting there are great. Nonetheless, this is what Kant tries to show in the deduction.

To show reason's practical nature, its ability to determine itself freely, Kant limits the proof of freedom by linking it to morality. But morality is only a particular type of autonomy, only a particular employment of the use of practical reason.⁴ Nonetheless, since Kant treats morality and autonomy as the same for much of the *Groundwork*, it is worth briefly examining the particularities of moral insight. Moral insight, as Dieter Henrich characterizes it, is knowledge of the good. But this knowledge is not simple theoretical knowledge, since it is unmediated. The good is always evidently good and needs no external justification. But the good is

³ Gerold Prauss. *Kant über Freiheit und Autonomie*. Frankfurt: Klostermann, 1983. 12-15.

⁴ What, exactly, this means, is in a sense the purpose of the dissertation in its entirety. For what is at stake in the uncoupling of ethical relations from universality makes up the story of much of the critique of Kant's rigorism. In short, the later authors to be considered below have held that reflection need not be taken to be universal so much as universalizing. This means that reflection reflects from its particular position in history with an eye toward universality by not necessarily thinking that it has (or ever can) achieved true moral certainty of the type Kant's moral theory seems to imply.

also normative. Knowing the good is to feel bound by it. Thus all immoral action involves turning one's back on the good, at least in the moment of action.

It is this immediacy of the good in moral action linked to the fact that while knowledge may be true or false, action never is. That causes the apparent difficulty in linking theoretical knowledge with practical knowledge. Evidently, practical and theoretical knowledge, even if fundamentally the same, constitute quite different ways of comportment in the world. In the deduction, Kant tries to show how we can get from theoretical knowledge to practical action through the notion of autonomous reflection.

I see the fundamental task to be accomplished by Kant's ethics as the 'proof' (through deduction or otherwise) that there is such a thing as practical reason or autonomy. I will be looking at three attempts to show this.

Kant's Deduction

Kant's deduction proceeds in three basic steps. The first is to show that rational beings have autonomy, the second is to show that we are rational beings and thus autonomous, and the third step is to show that this autonomy requires us to be moral. In this chapter, we will be concerned mainly with steps two and three. The problem here is that once Kant goes beyond autonomy to morality, he must disregard nature as a determining factor in human will and create the heteronomy/autonomy distinction which does not hold. Kant is wrong to deny

nature in his moral theory. This is a defect that he tries to make up for in his political philosophy.

A word about the reconstruction: I will be basing my account chiefly on Allison's interpretation in his *Kant's Theory of Freedom* as this account provides the most convincing version of Kant's argument (including its limitation). I will be adding the interpretations of others as I go. Initially my interest will be to provide a clear account of this most difficult of sections. Criticisms and the reason why Kant abandoned the deduction, will be dealt with after the argument has been reconstructed.

The Analytic Thesis and Reciprocity Thesis

Kant begins Section 1, entitled "The concept of freedom is the key to the explanation of the autonomy of the will"⁵, with the distinction between two types of freedom. Negative freedom, as has been established in the third antinomy of *Critique of Pure Reason*, is limited to the ordering of apperception under concepts which arise outside the subject and merely affect it. In this case, the subject might be purely determined by nature or god and still be able to order concepts. The positive concept of freedom involves the will giving itself causal laws.

Kant wants to show that the positive notion of freedom follows from the negative notion, that reflection occurs under principles. Kant views the

connection between these two terms as analytic: “If, therefore, freedom of the will is presupposed, morality together with its principle follows from it by mere analysis of its concept.”⁶ This is followed by the claim: “But the principle of morality— that an absolutely good will is that whose maxim can always contain itself regarded as a universal law— is nevertheless always a synthetic proposition; for, by analysis of the concept of an absolutely good will that property of its maxim cannot be discovered.”⁷

These two sentences prefigure the argument almost in its entirety. The argument must thus proceed as follows: The Analycity Thesis states that a fully rational will always acts autonomously, i. e. under its own laws. Freedom requires rationality, self-determination. Then it must be shown, in Section Two of *Groundwork* III, that rational wills act under the idea of freedom. Section Three shows that humans are rational beings, from which it follows that we are free. Section Four shows that rationality implies the notion of the categorical imperative, that is, that the moral law has validity for us. Morality, however, is a synthetic proposition and thus requires a deduction. Allison argues that what is being deduced in GMS III is not freedom itself, but morality (the synthetic

⁵ “Der Begriff der Freiheit ist der Schlüssel zur Erklärung der Autonomie des Willens” GMS 446-47

⁶ “Wenn also Freiheit des Willens vorausgesetzt wird, so folgt die Sittlichkeit samt ihrem Prinzip daraus durch bloße Zergliederung ihres Begriffs.“ GMS 447.

⁷ “Indessen ist das letztere [Sittlichkeit] doch immer ein synthetischer Satz: ein schlechterdings guter Wille ist derjenige, dessen Maxime jederzeit sich selbst, als allgemeines Gesetz betrachtet, in sich enthalten kann, denn durch Zergliederung des Begriffs von einem schlechthin guten Willen Kann jene Eigenschaft der Maxime nicht gefunden werden.“ GMS 447

proposition). Once morality is established, freedom can be derived from it. It is only in this circuitous way that Kant can move from negative freedom to positive freedom. In a way, it might be said that morality is only a stepping stone on the way there.

The Reciprocity Thesis

What Allison calls the Reciprocity Thesis is an interpretation of the Analytic Thesis. Positive freedom or self-determination can be interpreted in two ways; as practical freedom or as transcendental freedom. Practical freedom, the weaker thesis, implies only that a rational agent would be able to determine the means to an end. This is simply the ability to reflect on a task and choose one means to achieving it over another. Such reflection implies nothing about the ultimate source of the principle according to which we seek to accomplish the given end. This type of freedom is compatible with strong determinism since we might be motivated by genetic determinations all the way down. Allison argues that Kant rejects this weaker thesis.

To prove positive freedom, freedom must be taken in the strong, transcendental sense. A transcendently free agent adopts her principles because they are rational rather than because they are expedient (and thus informed by some inclination). Since they are adopted by the agent freely, it must be possible to give reasons for their adoption. Since there is no source higher than transcendental freedom from which the reasons could derive and which could guarantee their adequacy, this type of reason giving is exclusively

autonomous. Only conformity to an unconditional practical law could provide the required ultimate justification and autonomy.

We must now show that conformity to the practical law is both a sufficient and necessary condition of the justification of maxims. Allison argues that it is easy enough to show that transcendental freedom/self-determination is a sufficient condition for ultimate reason giving: something that is justified (that is, a sufficient reason) to all others surely is justified (a sufficient reason) to me.

But we must also show that transcendental freedom is necessary. For if we cannot show this, it might still mean that some other, universal but heteronomous source (like god), could be determining us to give the type of universal reasons we give. The problem is that the maxim only seems justified if it already takes the form of morality, i. e. is done from duty. But it cannot be that only that which is done from duty is self-determined. Allison thinks we can overcome this difficulty by means of the distinction between permissible and obligatory. We could thus say that the maxim must be justified only as *at least* permissible, i. e. at least not contradicting the self-determination of all others around me.

Since we are not concerned with appetitive beings here, we are not dealing with restrictions on desires and inclinations. We are dealing with what is permissible as such, i. e. rationally justifiable under any circumstances. But this is precisely what Kant understands by the unconditional practical law.

Having avoided pegging self-determination to duty alone, we still have a neutral conception of practical reason as self-determination or autonomy. We

merely have a universally valid justification which is empty of content. There are no moral laws which can regulate our conduct other than reflection itself.

Autonomy is to act under practical reason. Transcendental freedom means acting under the practical law.

Allison thinks that Kant has thus shown that practical freedom will not suffice to give a general justification for a maxim, but transcendental freedom will. This is because practical freedom is ends oriented while transcendental freedom is not. The key difference is retained in the idea of permissibility which, in this context, is not a moral conception (though duty or obligation would have been). So far we have shown that rational agents must be thought of as transcendently free. We have not determined this freedom with regard to any required types of action. Thus far we also do not know why transcendently free agents conform to the pure practical law.

The next step must tell us something about the unconditionally practical law's connection to the moral law, for it is through morality that Kant wants to show that freedom exists. The key move will involve some notion of intentionality which will reveal our connection to the moral law. In other words, Kant will have to show that conforming to the practical law is the same thing as conforming to the practical law *because* it is the practical law. We must thus connect what appears to be a 'fact' about rational free agents (their conformity to the practical law) with the intentional structure behind this 'fact' (why they do so). Allison argues that to say that "conformity to universal law must be the reason for

adopting a maxim is just to say that [the law's] merely legislative form must provide the reason [...], the 'determining ground' of the will."⁸

Transcendental freedom is the missing link since transcendental freedom as self-determination means that one's reasons are normative for oneself in an unconditional way. This, and here comes the problematic step in the argument, is because transcendental freedom is not connected to any desires or physical impediments. For transcendently free agents, actions and reasons for actions are one and the same. Practical freedom, by contrast, does not have unconditional norms because it is (at least at times) externally determined.⁹

But this now only establishes the categorical imperative for rational/transcendentally free beings (if, at this level, we can even speak of imperatives), but not yet for humans. Allison has thus established that Kant starts from a thick notion of transcendental freedom rather than the thin one of rational agency. Self-determination, in other words, is a normative act. We can now look back and see why Kant thinks that, for transcendently free rational agents at least, it is analytic that morality follows from the analysis of freedom. It follows because, for purely rational agents, to be self-determining, to have reasons for actions, just means acting on these reasons. And this means simply taking one's reason's to be motivating, which Kant believes to be an analytic claim.

⁸ Allison. *Kant's Theory of Freedom*.

⁹ In a sense, this is a precursor to the internalism argument we will see in chapter 3. To have a reason for an action is thus to act. Nothing is a reason unless it can motivate an agent.

The second part of the 'proof' of freedom concerns the question of whether humans have a rational will. Accordingly the next section, Section Two, will argue that rational wills cannot help but act under the idea of freedom, and that all rational beings with a will must be free.

The Preparatory Argument and the Argument against Fatalism

Allison argues that since the moral law is supposed to apply to us qua rational beings, and since (by the Reciprocity Thesis) it is supposed to be derived solely from the property of freedom, freedom must itself be presupposed as a universal property of rational beings with a will. But since, *per the Critique of Pure Reason*, freedom cannot be proven theoretically, the next best strategy is to argue from the weaker thesis that freedom must necessarily be presupposed, and this is just what Kant seems to do. The so called preparatory argument makes the point in two steps:

(1) "I say now: every being that cannot act otherwise than *under the idea of freedom* is just because of that really free in a practical respect [...]."¹⁰

(2) "Now I assert that to every rational being having a will we must necessarily lend the idea of freedom also, under which alone he acts."¹¹

¹⁰ "Ich sage nun: Ein jedes Wesen, das nicht anders als *unter der Idee der Freiheit* handeln kann, ist eben darum in praktischer Rücksicht wirklich frei." GMS 448.

¹¹ "Nun behaupte ich: dass wir jedem vernünftigen Wesen, das einen Willen hat, notwendig auch die Idee der Freiheit leihen müssen, unter der es allein handelt." GMS 448

The cash value of being free 'from a practical point of view' is that one is subject to whatever laws one would be subject to if, *per impossibile*, it could be proven (theoretically) that one is free. As Ameriks remarks, this means that a proof of our freedom 'in a theoretical respect' would add nothing about the truth of our freedom or its relevant practical consequences.¹²

One would expect Kant to go on arguing as follows (Allison calls this the extended argument):

"(3) All laws 'inseparably bound up with freedom' are valid for every being with reason and will.

(4) But the Reciprocity Thesis establishes that the moral law is 'inseparably bound up with freedom'.

(5) Therefore, the moral law is valid for every being with reason and will.

(6) Since beings such as ourselves have reason and will, the moral law is valid for us.

(7) Since we do not necessarily follow the dictates of the law (these dictates being 'objective necessity' but 'subjective contingency'), the law for us takes the form of a categorical imperative, that is, we are rationally constrained, although not necessarily causally necessitated, to obey it."¹³

The problem with this line of argument is that, as Henrich points out, Kant has not yet established that humans have a will, i.e. that they have a practical

¹² Karl Ameriks. "Kant's Deduction of Morality and Freedom". 170.

¹³ Allison. *Kant's Theory of Freedom*. 216.

capacity at all.¹⁴ Here Kant identifies practical reason with will. This invalidates points (6) and thus the entire argument.

Another perspective on the argument is given by Allan Wood who argues that the preparatory argument presents us with a sort of analogy of between rational norms and morality. He claims that Kant needs to argue that the capacity we ascribe to ourselves in regarding ourselves as subjects of moral obligation is of exactly the same *kind* as the one we ascribe to ourselves in thinking of ourselves as judging according to rational norms, therefore if we cannot intelligibly doubt that we have such a capacity in one case, we have no good ground for doubting that we have it in the other. That this claim is legitimate can be seen in the fact that we do not accept logical rules only temporarily. We follow logical rules because it is unconditionally necessary to do so in order to preserve the truth of our judgments. All those who think of themselves as making rational judgments already presuppose something that commits them to morality.

The point of the preparatory argument, as these commentators point out, is that thinking of yourself as free is, in a practical sense, already to be free. This is why a theoretical proof cannot make us more free, nor can a proof that we are in some sense determined take away from our freedom. Freedom is a first personal experience and thus thoroughly practical.¹⁵

¹⁴ Dieter Henrich. "Die Deduktion des Sittengesetzes". 64-70.

The Hidden Circle

Next Kant raises the problem of a circle in his argument.¹⁶ This problem is neatly formulated by Ameriks:

If we *define* a rational will as a will of its own in the sense that it is to obey only self-legislated principles, principles which according to Kant can be only of a universalizable type, then subjection to morality can follow from freedom of will. But the entailment involves a kind of circle, for the free rational will that is the premise leads to morality only because its freedom is understood not as merely (negative) transcendental freedom, but as autonomy.¹⁷

The problem, as stated by Ameriks, is that Kant simply has not yet established the positive concept of freedom needed for morality. He has established only that rational agents must take themselves to be free.

The circle has been much debated in the literature, but Allison's deflationary approach to the problem seems a good one. According to Allison, the problem of the circle is only apparently a problem for Kant since Kant has not actually made the argument that would trap him in the circle. At this point, the argument is only hypothetical: "If one regards oneself as a rational agent, then one must also regard oneself as free and therefore as standing under the moral

¹⁵ We should note, however, that putting things in this way already foreshadows the central problem for Kant, which is that freedom might turn out to be too 'first personal', that is, it might only apply to me.

¹⁶ GMS 450.

¹⁷ Ameriks. "Kant's Deduction of Freedom and Morality". 172.

law.”¹⁸ But Kant has not yet made the extended argument (points 3-7 above). If he had done so, however, without having established our rational nature (as he has not), he would have been guilty of a circular argument, i.e. assuming the idea of freedom only because of the moral law (rational will) in order subsequently to infer the moral law in its turn from freedom.¹⁹

This point serves again to emphasize the care Kant is taking in making the argument for freedom. It also helps to set up the problem of the actual deduction of morality and its implication of freedom.

The Two Standpoints Argument

The circle can only be resolved if Kant can show that we are rational agents with wills. Kant does this with the help of the two standpoints argument. Again we follow Allison. The two standpoints argument claims that our membership in the intelligible world provides the needed nonmoral premise from which our freedom and (by the Reciprocity Thesis) our subjection to the moral law can be derived in a manner which is neither question-begging nor circular manner. The argument must show that, qua members of the intelligible world, we are warranted in assuming that we really are rational agents and, as such, really subject to the moral law. This is also meant to show, then, that we have practical reason, i.e. the ability to determine ourselves through the practical law. This is

¹⁸ Allison. *Kant's Theory of Freedom*. 221.

¹⁹ GMS 453.

the most controversial step in the deduction and the one that ultimately drove Kant to abandon the deduction.

The key is now to show that we humans are members of the intelligible world. We thus need two more arguments: (1) an argument that, without appealing to either the moral law or presupposing freedom, establishes our entitlement to regard ourselves as member of the intelligible world, even if only from a 'point of view' or 'standpoint'; and (2) an argument showing that, qua members of the intelligible world, we are justified in regarding ourselves as rational agents, that is, as rational beings with wills. Given this, Allison argues, the validity of the moral law for us would follow by Reciprocity Thesis.²⁰ That is to say, Kant must show both that we are free and that this freedom is able actually to influence our empirical nature through our will.

The argument proceeds from the analytic claim earlier that we act on reasons. Then Kant adds that as cognitive beings we are conscious of capacities in virtue of which we cannot conceive ourselves merely as sensibly conditioned. That is, we are conscious of having reasons for our actions. So, knowing that we have a will must now be linked with membership in the noumenal world in which reasons are always acted upon. The problem is that mere consciousness of having a will is not enough to establish that we actually have a will. (If it were, there would be no problem of the circle and no recourse to the intelligible world would be needed.)

²⁰ Allison. *Kant's Theory of Freedom*.

The crux of the argument is that if consciousness of the will is also an activity of the will (in the intelligible world), this evades the danger that the will is merely illusory or determined from some outside source, (since transcendental idealism rules out any empirical effects on the intelligible world). From the two different perspectives, the will is thus real both in the intelligible and the sensible sense. It is both passive with regard to the object's nature and active in changing this nature.

This point comes down to the notion that self-consciousness, which might be deceived, cannot be deceived in its practical employment because practical employment is already an action within the world. There is thus a crucial difference between thinking oneself free and acting on such a thought. You might think you are free and be deceived, but if you think you are free and act as if you are free, you have just acted freely. And if you have just acted freely from *your* standpoint, no one can say that you have not acted freely from an 'objective' standpoint.

The Categorical Imperative

Thus far Kant has established that we (humans) may think of ourselves as both possessing a will and as being able to use it freely (in an incompatibilist and transcendental sense). He has not yet established that we have any reason to do so, or, to put it differently, that pure practical reason can actually motivate us and is therefore normative for us. What, in other words, is the relation between the 'is' of our self-determination, and the 'ought' of morality? This further step must be

accomplished in the section entitled “How is the categorical imperative possible?”²¹ Again we turn to Allison. This step presupposes that the moral law is the law according to which intelligible beings act, and that it is also the distinction between governance in intelligible and sensible worlds. Kant first notes that if we were purely intelligible beings, all of our actions would be autonomous; if we were solely part of the sensible world, however, all our actions would be governed by desires and inclinations.

The deduction itself comes at *Groundwork* 453-4. The nerve of this argument lies in the complex claim that since the intelligible world is (in general) the ground of the sensible world and its laws, and since the will as a member of the intelligible world is subject to its laws, the will must also be conceived of as grounded in the intelligible world and its laws. From this it supposedly follows that sensibly affected and therefore phenomenal beings such as ourselves, who are likewise conscious of possessing a will, experience the law stemming from this noumenal will as an unconditional command, that is, a categorical imperative addressed to them in their phenomenal nature.²²

It is worth pausing here to remark that this argument is really the crux of Kant’s theory of normativity. What makes a reason powerful enough to motivate us, that is, to beat out competing inclinations, is that its source stems from our noumenal or rational nature. What makes rational reflection motivating, in other words, is its purity, its *separateness* from the world in which the human agent

²¹ “Wie ist ein kategorischer Imperativ möglich? GMS 453-56

²² Allison. *Kant’s Theory of Freedom*. 225.

must act. The problem with this argument is that sensible creatures must be addressed by the moral law in the right way, i.e. they must be affected but cannot be necessitated. The separateness from nature which gives reason its authority also threatens to keep reason from affecting the natural world of inclination. Normativity must retain its motivational character and must thus appear in the form of a command. But it is not clear how this might occur.²³

Kant tries to confront this problem by arguing that the moral law is synthetic. Thus the synthetic a priori proposition (“an absolutely good will is that whose maxim can always contain itself regarded as a universal law”²⁴) has already been established in the deduction of the moral law, not merely at the introduction of the categorical imperative. The derivation of this synthetic a priori is only possible on the basis of the positive concept of freedom, which furnishes the third term enabling the link between the concept of the absolutely good will and the moral requirement. The will cannot be considered moral unless it is free in a transcendental sense. If freedom were of a lesser degree, we could not rule

²³ Allison argues that, with the benefit of hindsight, what Kant seems to need is his later distinction between *Wille* and *Willkür*, where pure *Wille* (or pure practical reason) confronts the sensibly affected (yet free) *Willkür* as an unconditional demand, one to which our needs as sensible beings must be subordinate. Allison’s argument is a version of the argument Kaulbach will make on a broader scale, namely that in order to unify ‘wild’ nature and reason, we really need a third term, the nature in the true self. But while Allison bases his argument on a reading of *Religion*, Kaulbach believes the point can be made more generally.

²⁴ “Ein schlechterdings guter Wille ist derjenige, dessen Maxime jederzeit sich selbst, als allgemeines Gesetz betrachtet, in sich enthalten kann.” GMS 447

out being influenced by an external force. Morality requires transcendental or incompatibilist freedom.

Finally, once one has established the status of the moral law as the principle of autonomy on which rational agents would necessarily act if reason were fully in control, there is no further difficulty in claiming that it is also the principle on which rational agents *ought* to act should they be tempted to do otherwise. This point is a consequence of the two stand point argument under which the is/ought distinction is a distinction merely of perspective and not of fact or ontology. So, by reducing the distinction between our phenomenal and intelligible nature to a matter of perspective, Kant tries to show that we can be affected by both. Our noumenal nature's superior authority, however, means that when it raises its voice in command, we *ought* to heed it and not our natural inclinations.

This point, however, needs further clarification. We can now turn to Dieter Schönecker and Allen Wood's analysis of this issue of transition from autonomy/freedom to morality. In doing so, we should keep in mind that what is being sought is an answer to how, in fact, we are to understand Kant's claim that the categorical imperative is merely the concept of the moral law under the conditions of a finite will, as Otfried Höffe has nicely put it.²⁵ To put it differently, how can the two standpoint argument maintain the ontological equality of both perspectives while maintaining the motivational (and therefore affective) superiority of the intelligible?

We saw earlier, in Allison's defense of the Reciprocity Thesis, that Kant was able to show that for a rational being (existing in the noumenal realm) there was strictly speaking no distinction between being free and being moral. To act autonomously simply *is* to act according to the categorical imperative, i.e. to treat people as ends and never as means only. Of course, for purely rational creatures, there can be no imperative properly speaking since what they want to do and what they should do are the same. This is not so for humans. Humans are impeded in their rational willing by desires and habits. What Kant must thus show is that the moral law has a motivating effect on rational human beings.

According to Schönecker and Wood the answer to this question is that I *should* act morally because I *can* act morally. But this answer is only possible if it can be shown that being-able-to-act-morally is fundamentally a wanting-to-act-morally. It is Schönecker and Wood's hypothesis that Kant grounds the wanting-to-act-morally on the claim that the intelligible world has a superior ontological status. Schönecker and Wood thus conclude that, in a fundamental sense, for Kant the 'true self' is really conceived of as wanting what it ought to want rather than what it wants to want.²⁶ The true self, in other words, is the noumenal self. But to say that is, again, to have separated the noumenal self from the sensible self to such an extent that they seem to have no contact at all.

²⁵ See Otfried Höffe. *Immanuel Kant*. 181-85.

²⁶ Schönecker and Wood *Kants 'Grundlegung zur Metaphysik der Sitten'; Ein einführender Kommentar*. 200-202.

For Schönecker and Wood the superiority of moral autonomy and what makes it appear as an imperative in the sensible world, is moral autonomy's ontologically superior status. For it is only in the difference between the intelligible and the sensible world that such a thing as morality can even appear; it is only beings who are intelligible and sensible at once who must be motivated to act on one of their parts rather than on the other.

Stressing the point made above about ontological separateness, Schönecker and Wood conclude that this argument is illegitimate because Kant uses the distinction between the sensible and intelligible worlds in an ontological sense, but seeks an epistemological conclusion about motivation from it. It remains unclear how the ontological distinction between the sensible and intelligible world can be bridged such that the sensible world could be affected (in the appropriate way) by the intelligible world.²⁷ What is ultimately required here is an explanation of how there can be such a thing as a non-necessary necessitation. How can we be divided and yet unified in the way Kant describes?

As noted earlier, Kant's answer seems to reside in the difficult two standpoint argument according to which we are free in a practical sense if we think we are: "every being that cannot act otherwise than *under the idea of freedom* is just because of that really free in a practical respect".²⁸ Much depends on what Kant means by 'idea' here. 'Idea' here cannot be anything other than

²⁷ Schönecker and Wood *Kants 'Grundlegung zur Metaphysik der Sitte'; Ein einführender Kommentar*. 206. See also Allison's criticism of the same issue above.

that which powers the normative force of the categorical imperative. Since we are not purely rational, we can only have an idea of the noumenal law while living under the influence of nature. So, in a certain way, the two standpoint argument boils down to the question of our relation to nature both in ourselves and outside of ourselves. And this question becomes the question of whether we can, in fact, act autonomously at all, i.e. whether we have pure reason.

To put the question another way, the issue at stake is whether the idea of the moral law is a representation (of knowledge) or whether it is a practical rule of motivation. It must certainly be the latter since the idea must bring with it stronger authority than does epistemic knowledge, which is limited by our inability to represent the thing in itself. The authority of pure reason in its practical employment, however, must be completely self-authorizing and this means being motivating, a reason for action.

Annemarie Pieper presents an argument which might help us see the two standpoint argument in a new light. She argues that the Categorical Imperative, as based on Kant's distinction between the intelligible and the sensible worlds, does not constitute the naturalistic fallacy (as Karl-Heinz Itling maintains²⁹) in that it illicitly connects an 'is' with an 'ought' (where this distinction is taken as pertaining to different ontological statuses of motivations and facts). Rather, she argues, the distinction must be seen from the correct point of view and since the

²⁸ "Ein jedes Wesen, dass nicht anders als unter der Idee der Freiheit handeln kann, ist eben darum in praktischer Rücksicht wirklich frei." GMS 448

²⁹ Karl-Heinz Itling. "Der naturalistische Fehlschluß bei Kant".

empirical will cannot be the recipient of the categorical imperative, and the intelligible world cannot be seen as motivating the sensible world directly, Kant seeks to resolve the issue by placing it in the context of the theoretical and practical standpoints. She writes:

The categorical imperative does not *exist*; it has no way of being and exists just as little as the world of the understanding does. [...] The categorical imperative only comes into view from the perspective of practical reason: as a practical law, that is, as the normativity of norm generating laws whose ground is itself the highest normative principle, the principle of freedom.³⁰

The point is then that the categorical imperative is a purely practical law, unintelligible from the theoretical (empirical) perspective. Its basis, freedom, is equally unintelligible from the theoretical knowledge. This much we already know from the *Critique of Pure Reason*. But this still does not directly address the problem of the seeming ontological superiority of the noumenal realm over the phenomenal realm.

The interpretation of Kant's deduction I have given over the past sections concludes that Kant is unable to provide a justification for the normativity of the moral law for humans because it relies on an unjustified distinction between the phenomenal and noumenal natures of humans. The two standpoint argument

³⁰ "Der kategorische Imperativ *ist* nicht; er hat keine Seinsweise und existiert ebenso wenig wie die reine Verstandeswelt. [...] Der kategorische Imperativ kommt nur aus der Perspektive der praktischen Vernunft in den Blick: als praktisches Gesetz, d.h. als die Verbindlichkeit von Normen begründende Norm, deren Verbindlichkeit wiederum ihren Grund im höchsten normativen Prinzip,

Kant uses to justify this distinction states that because we can think of ourselves as free we are free, where thinking here means reflection in the practical sense of setting means to an end. Such means-ends thinking, however, is difficult to bring together with purity of motivation required by Kant's notion of moral autonomy. Kant thus seems unable to bridge the gap between our noumenal nature which thinks of itself as free and our sensible nature which sets ends and pursues them. Kant, in other words, is unable to show that our end is freedom itself because he has not shown that pure practical reason itself motivates our end setting practices.

The two interpretations of Kant that follow seek to construct alternate readings of Kant's phenomenal/noumenal distinction. Prauss pursues what I think will turn out to be a rather Hegelian approach in which the process of reflection itself takes center stage. Kaulbach pursues the question of reason and nature from the perspective of historical development. He thinks that inclinations can be made more autonomous through a sort of cultivation of the 'true self', which is autonomous.

Prauss: Pure Practical Reason and Morality

We now move to the second approach to Kant's notion of pure practical reason. Prauss contends that Kant gives up his initial insight into the necessary

im Freiheitsprinzip hat." Annemarie Pieper. "Wie ist ein kategorischer Imperative möglich?". 279.

opposition between nature and freedom by developing his concept of autonomy exclusively through morality. The concept of morality forces Kant to value the noumenal realm as good and the sensible realm as evil. This means that action and theoretical self-relation must remain external to one another. If Kant starts from the perspective that this theoretical self-relation is on the side of inclination, it will be impossible to prove that autonomy exists. But if Kant starts from this theoretical self-relation as rational, then it is not at all clear how reason can ever be practical.

As we saw at the outset of *Groundwork* III, Kant develops the concept of freedom out of our theoretical understanding of spontaneity of the mind. He argued that freedom is expressed in our ability to order the sensory impressions of the world. As Prauss glosses it, this means that subjectivity determines merely the “‘projection’ of objectivity for its knowledge and but not yet anything corresponding for its actions”.³¹ Thus, Prauss argues, negative freedom cannot lead to positive freedom. This means, however, also that Kant cannot bridge the gap between theoretical reason and practical reason. The problem, Prauss contends, lies in the way Kant conceives of the ‘eigentliche Selbst’, ‘the true self’. The problem with Kant’s construction of practical reason is that Kant does not have a concept of the practical at all and merely tries to explain practical reason by way of the natural, that is, desires and habits. But, as previously defined, the

(my translation)

practical is for Kant “everything that is made possible through freedom” which excludes precisely this kind of heteronomy.³²

It is this problem which forces Kant to determine practical reason or will as a type of self-relation of the understanding or reason to its desires. This relation of desires is called a will precisely because of its relation to reason. But Kant still has difficulties distinguishing this type of self-relation from theoretical reason. As the example of the technical-practical imperative makes clear, theoretical reason is still of central importance in this conception. That is to say, the hypothetical imperative is still a theoretical means-ends setting capacity determined by the relations of empirical objects to the subject and is not in itself practical. Even here it is not clear what determines the will to act at all. Only the intentional structure is clear, given by theoretical reason, but practical reason seems not to have a place.

Since it must exclude the natural (inclinations and desires), the self-relation Kant has constructed appears to be based on theoretical reason. What I have explained in this kind of reasoning is merely a type of rational construction of interaction with the world based on our epistemological relations. Kant has simply remade self-consciousness as self-determination. The problem, Prauss contends, is that Kant has to claim that this type of self-consciousness of objects

³¹ Gerold Prauss. *Kant über Freiheit und Autonomie*.120. “Entwurf von Objektivität für ihr Erkennen und damit offenbar noch nicht auch schon etwas Entsprechendes für ihr Handeln”. (my translation)

³² Kant. *Critique of Pure Reason*. “alles, was durch Freiheit möglich ist” A800/B828

outside the subject is also a practical relation. The true self would thus be both theoretical and practical at the same time.

But this leads Kant to the paradox discussed earlier: either the self *is* determined by the natural world, in which case autonomy cannot be distinguished from heteronomy, or the self is reason, in which case it becomes hard to see what is practical about it at all (given the exclusion of the natural/practical world).

Practical Reason and Autonomy

According to Prauss, however, Kant makes one further attempt at grasping the nature of the practical by defining the practical as a self-relation of will and getting at autonomy in this way. Autonomy is legislation which is particular to oneself— ‘eigen’. In this sense, having a will means having a will of one’s own, a particular relation to oneself. Animals, accordingly, do not have a will. Prauss writes: “That humans possess something like a will at all [...] stems from the fact that they originally possess the will as a type of practical self-relation, a ‘will for oneself’ or a ‘willing of oneself’”.³³

It is only through this self-relation that the will can even be conceived as wanting anything for itself or others. And it is only this way of seeing the question

³³ Prauss. *Kant über Freiheit und Autonomie*. “Dass der Mensch überhaupt dergleichen wie einen Willen hat, liegt [...] ausschließlich daran, dass er ihn ursprünglich als ein praktisches Selbstverhältnis besitzt, als einen ‘Willen zu sich selbst’, oder ein ‘Wollen seiner selbst’.” 132. (my translation)

of autonomy that can make intelligible the possessive meaning of the 'own' in 'own will', or 'own act'. One's desires, however, remain external and can never be one's own in the reflective *and* possessive sense. Desires may be mine in a reflective sense, but they are also always causalities of nature and therefore not really 'mine' all the way down.

Kant equates this 'ownness' of will with autonomy and hence with being an end in oneself. The essential self-related act of willing is thus a reaction of the self to itself in which the self is its own end. In this way Kant can separate what has a relative value from that which has absolute value. If 'means' are to have any value at all they can only have value relative to something which has absolute value, namely an end in itself. Having a will of one's own is like being an end in oneself in the sense that both are self-directed. Having a will means willing oneself and being an end in oneself means being one's own end. "Whatever else he might want, man is this willing only because in willing whatever else he wills, he always fundamentally wills himself".³⁴

We might take Prauss' argument about willing oneself as the missing analogue of the apperception thesis in which we know ourselves knowing ourselves. This is the fundamental self-relation of German Idealism. The idea that the self wills itself in everything it wills means that we have reversed the direction of the dependency between the understanding and reason. Practical

³⁴ Prauss. *Kant über Freiheit und Autonomie*. "Was auch immer er wollen mag, so ist der Mensch doch eben dieses Wollen jeweils nur dadurch, dass er in allem, was er will, grundsätzlich einmal sich selber will." 135-36. (my translation)

and theoretical reason are now coeval in the sense that both are required for human action to be intelligible.

Thus we can see that, initially at least, Kant's practical philosophy defines the human not merely as theoretical self-understanding but also in a practical but morally neutral way. Humans have inner absolute worth *because* they are will themselves and so are 'of value to themselves'. Thus Prauss concludes:

In order to have a fundamental consciousness of and a will toward himself in a neutral sense, which would then also be both a theoretical as well as a practical self-relation, the agent would also have to be autonomous in a neutral sense; and it would be from this autonomy that Kant could then derive moral autonomy.³⁵

Thus only humans as ends in themselves can form the basis for the principle of morality, and not morality itself.³⁶

In a sense, Prauss' two arguments make two opposing points which flank Kant's argument. While the first argument we saw contends that Kant's view of morality is illegitimate because it amounts to a division of the intelligible will into

³⁵ Prauss. *Kant über Freiheit und Autonomie*. "Und dem neutralen Sinn gemäß, in dem der Mensch danach ursprünglich ein Bewusstsein von und Wille zu sich selbst und damit Freiheit eines theoretischen wie praktischen Selbstverhältnisses wäre, müsste er zunächst auch in neutralem Sinn schon Autonomie sein, aus welcher Kant so etwas wie moralische Autonomie dann überhaupt erst abzuleiten hätte." 137-38. (my translation)

³⁶ Schönecker and Wood make a similar criticism. They argue that even if one accepts Kant's distinction between things in themselves and appearances, it is far from clear how the fact that humans are part of both realms allows us to infer that desires must be subject to reason (given the invalidity of the ontological thesis). Kant's ultimate answer, they contend, is that humans are ends in themselves and thus valuable. But rather than explain what this might mean (as Prauss attempts to do) Kant gives only an implausible ontological answer to this question. Schönecker and Wood *Kants 'Grundlegung zur Metaphysik der Sitte'; Ein einführender Kommentar*. 204-205.

good and the sensible realm into evil, the second argument about the self-relation of willing, states that the notion of an autonomous self-willing subject does not even amount to an explanation of the moral law because being an end in oneself and being a means to an end are not stable categories. Both of these arguments suggest that Kant has not been able to overcome the ambiguity in his notions of heteronomy and autonomy.

Prauss' criticism is thus fundamentally concerned with Kant's distinction between intelligible and sensible. The best way of developing a satisfactory interpretation of this issue is to develop a concept of autonomy, which Prauss believes can be found in the notion of a self-relation of willing. But this too does not provide for a satisfactory answer to the ethical imperative which Kant claims stems from the noumenal realm.

Because the question of this equal relation between nature and morality seems to involve the metaphysics of the noumenal realm, Anglo-American commentators have tended to shy away from this part of Kant's ethics, preferring to ground their interpretations on the notion of humans as ends in themselves.³⁷ Prauss' critique of Kant, however, hinges on taking seriously the sensible-intelligible distinction since it also takes seriously the apparent exclusion of the sensible realm from Kant's moral theory.

³⁷ Allison, of course, is a notable exception.

Kaulbach: The Historical Perspective on Autonomy

Friedrich Kaulbach's interpretation of Kant takes up the problem of autonomy by seeking a historical rather than theoretical resolution to the problem of moral judgment. It takes up the problem of the noumenal realm by providing a 'dialectical' interpretation of Kant's ethics. The central claim, which echoes the point at which we left off our discussion of Prauss, is that the subject can effect a reorientation of the emotions through reason on the basis of autonomy. This reorientation is meant to explain how reason (or the noumenal) is able to affect our finite nature. Kaulbach also raises the speculative point that it is inadmissible to cut off nature from our practical deliberation. Nature, through history, must be given a place if the employment of practical reason is to be normative for us.

The reorientation of the self assumes not two but three parties involved. Kaulbach believes that humans have a fundamental nature, something like a substratum affected by our 'wild' nature and our reason. Thus the true self (Kant's 'eigentliches Selbst') is able to create a type of consensus between our emotions and our reason which means that our egoistic feelings can be brought into unity with a furthering of general human life. Kaulbach writes:

The practical identity which arises in this manner is that of a self in accord with itself to the extent that its free reason (*Vernunft*) also permeates the reason of initially foreign beings. [...] Insofar as I become myself through the history of my actions and to the extent that my own 'nature' is permeated by the said 'change of disposition' (*Um-stimmung*), I have the right to call myself *free*.³⁸

³⁸ Friedrich Kaulbach. *Das Prinzip Handlung in der Philosophie Kants*. 168. „Die praktische Identität, die so entsteht, ist Übereinstimmung des Selbst mit sich, sofern seine freie Vernünftigkeit auch die der Vernunft zunächst fremden Natur durchdringt. [...] Sofern ich durch

We become this true or rational self through a historical process: “The categorical imperative asks agents to constitute their consciousness through practical thinking in such a way that it becomes possible to *recognize* what our duty is and then to *decide* to act on what duty demands.”³⁹ The maxim itself is thus the measure or indication of the reasoning subject’s degree of rationality.

This reorientation allows me to test the universality of my impulses against the prescription of the categorical imperative. The maxim brings together theoretical and practical reason because anything it seeks to implement will be both empirically determined and practically implemented. The subject must determine whether the maxim, which arises out of interest (and is thus normative in a broad sense) is in accord with the principles of living which the true self endorses (and is thus normative in the more specific rational sense of just). The maxim is a product of autonomous thinking if it meets the criterion of reason coinciding with itself and therefore autonomy. But this is only part of the question about how the categorical imperative can be normative for us.⁴⁰

die Geschichte meines Handelns Ich selbst werde, sofern auch meine ‚Natur‘ durch die beschriebene Um-stimmung meiner Triebshäre mit Vernunft durchdrungen ist, vermag ich mich für *Freiheit* zu beanspruchen.“ (my translation)

³⁹ Friedrich Kaulbach. *Das Prinzip Handlung in der Philosophie Kants*. 171. „Der kategorische Imperativ fordert den Handelnden auf, in sich auf dem Wege praktischen Denkens eine Verfassung seines Bewusstseins herzustellen, in der es ihm möglich ist, die ihm angemessene Pflicht (Maxime) zu *erkennen* und sich zugleich zu ihrer Verwirklichung zu *entscheiden*.“ (my translation).

⁴⁰ This is also as far as more recent approaches to the maxim problem have gotten. See Rüdiger Bittner. „Maximen“. (Akten des 4. Internationalen Kant-Kongresses. Mainz 6-10 April 1974. Teil

The more fundamental question to be answered is how theoretical and practical reason combine to move us to action. How is it, in other words, that we can be both subject to the laws of nature and at the same time makers of the 'laws of nature'? How are we to understand the dialectic of freedom and nature, of autonomy and heteronomy, of autonomy and influence? The reorientation Kaulbach introduced earlier relied on what I called the substratum of the 'true self'. But how can the objective domination of the self by nature be made to coincide with the subjective idea of freedom which we have through reason?

Kant's answer to this problem, according to Kaulbach, is that there is in effect no difference between the interest of reason and the interest of nature, but that both are simply interests. The difference is that the interest of reason is true to itself (and therefore autonomous) while the interest of 'wild' nature is pathological (and therefore heteronomous). But both work on the same subject. The interest of reason is superior to the interest of 'wild' nature because reason is at once both a source of practical argument (reason in the logical sense) and a cause of action (reason in its practical sense) while 'wild nature' is only the latter.

This distinction calls for a clarification since it seems that it might be begging the question to assert that reason is both practical and theoretical at

II.2 Hg. Gerhard Funke. pp. 485-98.), Höffe. „Kants kategorischer Imperativ als Kriterium der Sittlichkeit“. (*Ethik und Politik*. Frankfurt: Suhrkamp, 1979. pp. 84-119.), even Barbara Herman's seminal essay „The Practice of Moral Judgment“ (*The Practice of Moral Judgment*. Cambridge: Harvard UP, 1993. pp. 73-93.) seems not to inquire into the conditions under which maxims might be normative.

once and that practical reason is thus its own cause. In answering the question, Kaulbach argues that the two perspectives of theoretical and practical reason permit us to say that freedom is both self-verifying and self-actualizing.⁴¹ In reorienting itself to become more true to itself, the good will is seen as creating a 'second nature', a practical being of a second order which is not determined by natural inclinations but by freedom. Under the category of causation, this 'intelligible' order is a cause because it is self-determined. Respect for the law now appears as the limiting and quashing of the wild forces of nature in us. Practical reason thus transforms the meaning of egoistic self-love into a rational will which issues in 'rational self-love'. Again it must be emphasized that we are not speaking of limiting our nature itself, but of limiting the egoistic interests in our self. These two senses of nature must be kept separate.

What drives this reorientation and how can reason dominate nature in us? Kant conceives of theoretical knowledge not in terms of pure truth but in terms of a critique of the notion of truth as truth for us. Kaulbach argues that theoretical reason projects its practical application into the world in the form of the categorical imperative. This practical constellation is the perspective of freedom under which we see the world:

The practical constellation is produced by the subject in 'thinking-action'. Here the agent assumes the role of the practical being determined by reason, sets the meaning of good and evil and assigns the appropriate moral value to ends and means. These

⁴¹ That both verification and actualization are essential parts of any post-metaphysical or speculative philosophy will be a central concern of not only Hegel, but also Korsgaard. See chapters 4 and 5.

categories make it possible to think and speak about the purpose [*Gegenstand*] of action. This constellation has the character of freedom to the extent to which it is produced by the self-determining practical subject.⁴²

The practical constellation provides a perspective from which the epistemic understanding of the world (the 'is') already has a causal character (the 'ought') which means that not only is the moral law determined through knowledge of the world, but the moral law is also determining or normative for the subject. In this sense it can be said that freedom is both causality and cause.

This unity is forged by the resolution of the two standpoint argument in favor of reason in the dialectic of the *Critique of Practical Reason*. Kant's aim in the dialectic is to go beyond the question of moral 'knowledge' to the question of meaning (*Sinn*). According to Kaulbach, the discussion of meaning is also meant to show the how the elements in the series of (historical) maxims which testify to the reorientation of reason are related. This is the purpose of the argument about hope in *Critique of Practical Reason* and also in the historical writings. Hope is the (metaphysical) belief that reason will eventually be able to rule over 'wild' nature and that freedom will reign in the kingdom of ends. The notion of hope is

⁴² Kaulbach. *Das Prinzip Handlung in der Philosophie Kants*. „Praktische Konstellation [...] wird vom Subjekt in einer Denk-handlung hergestellt, in welcher ich mir die Rolle des praktischen, von der Vernunft bestimmten Seins gebe, in der ich auch die Sprachregelung über Gut und Böse festsetze und Zwecken und Mitteln den ihnen gebührenden moralischen Wert erteile. Auf der Basis dieser Kategorien ist Denken und Sprechen über die Gegenstände des Handelns möglich. Dieser Konstellation eignet der Charakter der Freiheit, sofern sie von dem sich selbst bestimmenden praktischen Subjekt hergestellt wird.“ 290. (my translation) This perspective

thus designed to hand the victory of the struggle between nature and reason to reason, for it is only reason which is capable of ultimately resolving the antinomy in favor of the idea of the kingdom of ends.

The metaphysics of hope is thus well suited to help us parse the world in terms of rational and non-rational objects, and this will further the goal of a final resolution to the struggle. As orientation or reorientation, the notion of freedom thus has concrete 'epistemic' ramifications for our actions, specifically in the sense that it permits us to think of freedom as causal in a more powerful way.

By conceiving of himself as a free subject in the practical constellation and thus making himself the true origin of his action, the agent allows the category of substance necessary for action to become a unifying principle, a principle through which the multitude of instances and moments which constitute the history of action become a unity of 'my' actions. Actions makes causality into a real concept.⁴³

From this Kaulbach concludes that, from the practical perspective, spontaneous action is not restricted to merely understanding (*begreifend*) reason but also to actualizing (*entwerfend*) reason.⁴⁴ Thus actions produce results which exhibit the mark of real freedom. "History appearing as the succession of action

seems also to be suggested by Kant's question "Was darf ich hoffen?" in the "Canon of Pure Reason" discussed in the introduction.

⁴³ Kaulbach. *Das Prinzip Handlung in der Philosophie Kants*. 298. „Indem sich der Handelnde als freie Substanz in der praktischen Konstellation prädiziert und sich dadurch die Wirklichkeit einer Ursache für Handeln gibt, lässt er die dabei einfließende Kategorie der Substanz zugleich auch als eininigendes Prinzip wirken, durch welches die Vielheit der Augenblicke und Momente, aus denen die Geschichte der Handlung besteht, zur Einheit eines einzigen Vollzuges jeweils ‚meiner‘ Handlung zusammengefasst wird. Kausalität wird durch Handeln zum realen Begriff.“ (my translation)

is thus not only a causal process but the 'externalization' of the free internal movement of the acting will".⁴⁵ This means that actions can actually be determined by the noumenal realm and hence that practical reason exists.

Kaulbach's interpretation both justifies Prauss' criticism of Kant's ethics and also tries to defend Kant. Kaulbach's argument can be taken to emphasize the necessity of developing a notion of autonomy in Kant which will then be the ground for a theory of morality. The division of human nature into 'wild' nature and 'reason' creates a three part structure rather than a dualistic structure usually associated with Kant, and thus suggests that the struggle between reason and nature can be carried out on the field of this third term. This substratum, as I have been calling it, assumes autonomy and conceives of the struggle taking place as not between heteronomy and autonomy but between heteronomy and morality.⁴⁶

The deduction in *Groundwork* III shows that autonomy is necessary if morality is to be possible. But the failure of the deduction also makes clear that

⁴⁴ In support, Kaulbach cites Kant. XVIII, Refl. 5612.

⁴⁵ Kaulbach. *Das Prinzip Handlung in der Philosophie Kants*. 318. „Die sich als erscheinende Geschichte vollziehende Handlung ist demnach nicht nur ein kausal beschriebener Prozess, sondern ‚Äußerung‘ einer freien inneren Bewegung des handelnden Willens.“ (my translation)

⁴⁶ In this sense, Kaulbach's interpretation, especially of this interpretative nexus has a certain affinity with Heidegger's notion of care as autonomy (see *Sein und Zeit*. Tübingen: Niemeyer, 1993. 196-97). Heidegger assumes care as an existential modality thus avoiding the whole autonomy/heteronomy debate. This allows him to interpret 'heteronomy' in the form of 'das Man'

autonomy is not the same as morality. Morality is rather an instance of autonomy which Kant uses to argue that autonomy is real. Thus the failure of the deduction leaves not only morality in question but also its foundation, autonomy. This shifts the burden of proof for Kant's ethics away from showing the existence of morality to showing the existence of autonomy.

Since, from the *Critique of Pure Reason*, we know that freedom or autonomy cannot be proven as such, the task of giving an explanation of autonomy must be accomplished in a different way. Kaulbach's defense of Kant's ethics thus seeks a different path. Kaulbach interprets the question of autonomy in a historical and materialist vein. The key interpretative move here is to distinguish between the theoretical understanding of the world as instantaneous and the practical understanding of morality as historical. The historical dimension is meant to provide room for the reorientation of the self's autonomy toward reason.

This historical interpretation of ethical decision-making can be faulted on two fronts. It takes autonomy for granted at the outset and does not succeed in providing a justification of it by the end of the argument. The argument thus replicates the structure of the *Groundwork* III deduction. Kaulbach substitutes a historical/ phenomenological explanation of morality for a deduction. He seeks to ground autonomy on morality in much the same way Kant did. But he fails because he, too, is unable to overcome the division between the sensible and

as a deficient mode of being rather than as a fundamental alternative, as Kant seems to have to. See also Prauss. *Kant über Freiheit und Autonomie*. 145.

the intelligible realms. As Hans Lenk has pointed out, Kaulbach tends to over-psychologize what Kaulbach calls the 'inner actions' motivated by freedom.⁴⁷ I add that the idea of freedom ultimately can only be turned into a causal power if we attribute psychological force to the intelligible realm. But this cannot be for two reasons: firstly, because Kaulbach does not succeed in truly collapsing the distinction between the intelligible and sensible realms and, secondly, because, to do so, would take away the forward thrust of the whole project of reason in Kant's philosophy, undermining the crucial historical premise. To actually unify reason and nature, if only psychologically, would impede any notion of hope which is designed to hold out a possible resolution but not to provide one.

Conclusion

Kant's failure to prove the existence of autonomy and morality comes down, I believe, to the failure of the two standpoint argument to be satisfactorily resolved. It is, at bottom, unclear how (even in Kaulbach's resourceful interpretation) the intelligible realm can affect the sensible one. But I believe that the reasons that drove Kant to postulate this distinction cannot very well be rejected either. For, as we have seen, the problem of autonomy and of practical reason is a question about the relation of the self and nature. To reject this distinction would be to fall back into the Aristotelian problem of having to equate interest and desire purely with nature and thus of having to reduce the concept of

⁴⁷ Hans Lenk. "Freies Handeln als Interpretationskonstrukt". in *Kant in der Diskussion der*

freedom to natural causes as well, or as Kant has it, to freedom on the turnspit. If we are to make sense of freedom we can only do so by opposing it to nature in some substantive way. This is a point of which both Allison and Prauss are keenly aware and is the reason why both insist on an incompatibilist notion of freedom in Kant. For to endorse a compatibilist idea of freedom, as many current commentators do, is to invalidate what Kant means by moral inquiry.

The two interpretations of Kant's moral philosophy I have offered have tried to go beyond the letter of Kant's text in order to reconstruct a position from within Kant which seeks to overcome the metaphysical residue of the intelligible/sensible division. I take these two interpretations to be paradigmatic of what such an overcoming of metaphysics in Kant would involve. On the one side, overcoming Kant's division of the world into intelligible and sensible would involve a stronger notion of autonomy, one that is more analogous to the apperception thesis which serves to justify Kant's deduction of the categories. As Prauss argues, what the apperception thesis has that the deduction of freedom lacks, is a strong notion of self-relatedness. For reason to be able to truly justify itself, reason would have to show that it does not rely on the exclusion of a central realm of human life.

I take Kaulbach's interpretation as complementary to Prauss' in the sense that Kaulbach argues that if autonomy is really to be understood as self-grounding and thus as self-related, such a self-relation will not only have to be a

Moderne. ed. Gerhard Schönrich and Yasushi Kato. Frankfurt: Suhrkamp, 1996. pp. 256-67.

mediation of intellect and nature, but a temporal mediation in which we pay close attention to the future-directed character of practical reason.

Both commentators, I want to argue, take it as an unacceptable consequence of Kant's critique of metaphysics that he eliminates altogether from his philosophy the interaction between world and subject in order to rein in the excesses of reason.⁴⁸

⁴⁸ I have not discussed Kant's other justification of reason, the fact of reason doctrine, which argues that the existence of freedom cannot be proven but must be assumed simply from the idea that we can imagine ourselves as free. There are two reasons for this. The first is that believing ourselves to be free rather than proving that we are so, while it alters the proof structure of Kant's moral theory, does not change what I have concentrated on as Kant's central thesis, namely the idea that in order to have a moral theory, the world must be divided into intelligible and sensible realms.

The second reason for not discussing the 'fact of reason' doctrine here is that I believe that what contribution it can make to our understanding of the self-grounding of reason, will better come into focus after we have dealt with the other option on the table for a grounding of morality: to learn from Kant's metaphysical faltering and thus to provide a moral theory based solely on a form of coherentism, as in the case of John Rawls, to whom I now turn.

CHAPTER II

RAWLS AND CONSTRUCTIVISM

The metaphysical problems that plagued Kant's deduction of morality in the *Groundwork* have seemed, to many twentieth century philosophers who wanted to retain much of Kant's moral philosophy, so great that these contemporary thinkers have abandoned the attempt to ground pure practical reason altogether. The question I mean to pursue in this chapter is whether a certain type of Kantian moral philosophy can get by without such a grounding. In Rawls one finds a writer who believes that much of Kant's ethical theory can be salvaged if one sidesteps the question of a metaphysical justification for morality and concentrates on the proceduralism necessary for universal justice.

Rawlsian constructivism is, as I hope to show, a worthy successor to Kant in the sense that it seeks to avoid the problems that have plagued generations of Kant interpreters— to find some way of making the categorical imperative 'work'. Rawls' strategy, by contrast, is to concentrate on the categorical imperative as a way of thinking about moral laws immanently, that is, as constantly articulated and enacted by the individual agent. For Rawls, the categorical imperative is just the mental process we engage in when we think about how to be just to other human beings. Rawls thus emphasizes respect for persons over moral psychology. Respect for persons entails that we treat others just as we want to

be treated by others and this simply means, not seeking special treatment for oneself. Respect, Rawls argues, should (and generally does) enter into every thought about others. This type of thinking is modeled in both of Rawls' justifications for the liberal political society: the original position and the reflective equilibrium. The categorical imperative is a way of thinking which enables such respect for others.

I will argue, however, that, compelling though Rawls' interpretation of Kant's ethical theory is, its aim of presenting a non-metaphysical interpretation is only partially successful. Rawls is successful in giving a non-metaphysical account of reflection through the reflective equilibrium— a process in which each agent reflects on her considered beliefs and also takes into account the beliefs of others. Absent a universal (and therefore 'metaphysical') notion of practical reason which underlies such reflection, however, there is no way of showing that the conclusions of individual reflection cohere in any socially meaningful way. Indeed, this absence of cohesion is the result of Rawls' failure to take concrete suffering into account. By building his theory on the possibility of coherence between individuals, Rawls has, I will argue, sidestepped the problem of the perspective of justice altogether.

In reconstructing Rawls' thought, I will present the argument regressively, starting from Rawls' conception of autonomy and working backwards, always asking for a justification for the previous level of argument, until at last we arrive at the reflective equilibrium which is supposed to underwrite the whole

conception of justice. The regressive reconstruction follows the argument Rawls gives in “Kantian Constructivism in Moral Theory”, if not in *A Theory of Justice*, and underlines the acknowledged debt Rawls owes to Kant. The regressive argument also affirms that, after all, Rawls wishes to give a Kantian style grounding to his project.

The Original Position and the Categorical Imperative

Rawls writes:

The original position may be viewed [...] as a procedural interpretation of Kant’s conception of autonomy and the categorical imperative within the framework of an empirical theory. The principles regulative of the kingdom of ends are those that would be chosen in this position, and the description of this situation enables us to explain the sense in which acting from these principles expresses our nature as free and equal rational persons.⁴⁹

For Rawls, the original position is a regulative principle and thus a way of adjudicating between conflicting desires and inclinations.⁵⁰ The agent in the original position must be both autonomous and motivated by her reflection. That is to say, the original position must yield universally acceptable principles (as in the categorical imperative) and it must ensure that these principles are

⁴⁹ John Rawls. *A Theory of Justice*. 226. Henceforth TJ when cited in the text.

⁵⁰ Rawls himself does not believe that Kant’s categorical imperative actually provides a particularly good way of determining a content of the moral law. This is what his own theory of justice is supposed to provide. Rawls. *Lectures on the History of Moral Philosophy*. 163.

acceptable to all. The former condition is modeled in the original position by bargaining and the latter is modeled by the veil of ignorance.

Let us look at autonomy first. Rawls introduces the veil of ignorance to hide the parties' particular social and natural circumstances. The parties are asked to design a society without the knowledge about where they will be placed in the society, or which beliefs, moral, political or religious they will have.⁵¹ All participants understand the basics of political affairs and economics and possess general knowledge. Thus they choose principles under which they are prepared to live, wherever they end up in society. The general social structure is just but blind to the particular inclinations of the agents. (TJ 118-19) Under the veil of ignorance, just as in Kantian autonomy, we have no personal or particular sense of the good. We seek only justice, the ability to enjoy our particular notion of the good once we determine what that is.

Rawls also argues that there is a parallel between rational choice theory and the categorical imperative. Rawls says that the original position is in the tradition of social contract theory. Like the categorical imperative, it provides a way of responding to a practical problem: what ought I do? Rawls' two principles of justice are simply the moral law under the conditions of a modern liberal

⁵¹ There has been considerable objection to the supposed neutrality made possible through the veil of ignorance. Onora O'Neill, for instance, notes that Rawls does not assume disinterest at all times during the original position process, but permits it with reference to the fate of future generations. "The Method of *A Theory of Justice*". These objections point to problems with the basic assumption of rational action. It is not clear whether we can ever be made neutral. See also the discussion of Sandel below.

society, yielding more specific versions of the universal prescriptive of respect as stated in the categorical imperative.

The parallel is this: the original position models the reflection of the individual on her goals with reference to the moral law as taking place between individuals under the veil of ignorance. What can be agreed on by all parties involved as the best way for each person to secure her ends in view of the alternatives available is, by definition, just and what is just, is what should be done. But rather than rely on what he believes is a dubious notion of the purity of practical reason, Rawls casts reflection as public deliberation between people under the veil of ignorance. (The reasons for the switch in perspective will be an issue below.)

This parallel, however, masks a point of divergence between rational deliberation as Rawls conceives of it and rational deliberation as Kant conceives of it. The difference concerns an ambiguity in Rawls' account of deliberation. For, as we saw in the last chapter, Kant's argument was that it was analytically true that the rational being would do what was rationally demanded of her by the categorical imperative. This was the weak autonomy thesis. The problem was showing that humans are rational beings— this was the strong autonomy thesis. The question to be examined here is to what extent Rawls' rejection of Kant's metaphysical division between sensible and intelligible worlds affects the justification of morality. Since we cannot yet give an answer to the question of whether Rawls has anything to say about the move from weak autonomy to strong autonomy, I suggest that we flag this problem by employing a distinction

Christine Korsgaard makes in her *Sources of Normativity*— the distinction between the categorical imperative and the moral law. The categorical imperative or CI-procedure is meant to denote rational deliberation itself, while the moral law is supposed to tell us what the range of such deliberation is. That is, all laws are universal, the question is whether the law ranges over my own life (in which case I might be an egoist) or over everyone, as Kant supposes.⁵² The question we must keep an eye on, in other words, is this: what sort of principle is involved in Rawls' notion of deliberation in both the original position and in the subsequent deliberation of the reflective equilibrium. Is it an agent internal rationality or an overarching intersubjectively rational law?

Let us return now to Rawls' reinterpretation of the categorical imperative as rational deliberation, the CI-procedure. The two principles of justice, as expressions of the moral law, are an equilibrium point. Rawls writes: "Equilibrium is the result of agreements freely struck between willing traders. For each person it is the best situation that he can reach by free exchange consistent with the right and the freedom of others to further their interests in the same way." (TJ 103) Elsewhere Rawls characterizes his theory of justice as partially a theory of rational choice.⁵³ We can thus say that the original position is supposed to provide a situation in which the participants can determine what they can expect to attain in terms of the advancement of their own ends given the ends of others.

⁵² Korsgaard. *The Sources of Normativity*. 99.

⁵³ TJ 15.

To put it another way, we could say that in the original position agents become clear about what they are maximally entitled to without invoking exceptions for themselves. No one, Rawls says, will take less than they think they can get.

Autonomy and the categorical imperative procedure are, of course, equivalents. Thus Rawls can say that the rational process of deliberation ensures that the particular conception of justice chosen is equivalent to saying that rational and disinterested deliberation had taken place (TJ 120). “The veil of ignorance makes possible a unanimous choice of a particular conception of justice.” (TJ 121)⁵⁴

We should note two points before we go on. Rawls has moved moral reflection from the first person perspective to public deliberation; from the ‘I’ to the ‘we’. At least *prima facie*, the original position is not supposed to be all in the mind of one individual. The second point follows from the first. By changing the perspective of reflection from the first person to the third person, Rawls has also changed the moral psychology involved in accepting the outcome of deliberation.

⁵⁴ Many have argued that a hypothetical agreement does not constitute a justification for the two principles chosen in the original position. See Thomas Nagel. “Rawls on Justice”. p. 6. Habermas. “Diskursethik— Notizen zu einem Begründungsprogramm”. p. 89. Along similar lines, R. M. Hare argues that agreement between the people in the original position merely gives us agreement on peoples’ opinions. These are thus merely a form of intuitionism, albeit cross examined. “Rawls’ *A Theory of Justice*”. pp. 83-84. See also Michael Sandel. *Liberalism and the Limits of Justice*. 109.

It is not clear that accepting the outcome of public deliberation has the same normative force as accepting the outcome of my own deliberation.⁵⁵

The Moral Character of the Ideal Agent

Underlying the original position and the application of the categorical imperative, however, is a conception of the moral character of the actors who reflect and thus abide by the moral law. In “Themes in Kant’s Moral Philosophy” Rawls interprets these agents as both ‘reasonable and rational’. Rawls uses these terms as a translation for Kant’s *vernünftig*, which includes both senses. The two terms mark the distinction Kant makes between the two types of practical reason, pure and empirical practical reason. The former is found in the categorical imperative while the latter is exemplified in the hypothetical imperative. Rawls notes that Kant’s conception of a person also marks the fact that, for him, the hypothetical imperative (empirical practical reason) is absolutely subjugated by the categorical imperative (pure practical reason).⁵⁶ This is to say that the person who engages in moral reflection subjugates his rationally conceived maxims to the moral law.

The same goes for Rawls’ agents in the original position. The original position too requires a certain conception of the person if it is to yield agreement

⁵⁵ In a way, this is the problem Rawls will have to address in *Political Liberalism* where he will have to show that we accept the results of the original position for reasons that are in a sense pure or moral rather than prudential.

⁵⁶ Rawls. “Themes in Kant’s Moral Philosophy”. 503-504.

at all; the people negotiating must not only be rational but they must also have a sense of justice, i. e. recognize that they are all equal. This, again, is the question of respect for persons. Thus “a society satisfying the principles of justice as fairness comes as close as a society can to being a voluntary scheme, for it meets the principles which free and equal persons would assent to under circumstances that are fair”.⁵⁷ This point presupposes that people must be free and equal in order even to engage in rational deliberation. They must be free in order to negotiate but they must also be equal in the sense of not believing that anyone deserves more than any other. I think Rawls believes that this captures the notion of ‘reasonable and rational’ in Kant, though this parallel is not made explicit until “Kantian Constructivism in Moral Theory”.

However, in *A Theory of Justice* Rawls already states that “the intuitive idea [of the two principles of justice] is that since everyone’s well-being depends upon a scheme of cooperation without which no one could have a satisfactory life, the division of advantages should be such as to draw forth the willing cooperation of everyone taking part in it [...]”.⁵⁸ Though Rawls has not yet explicitly made the argument for it, the claim is that the (instrumental) rationality, which is supposed to generate agreement through maximization in the original position, already includes some form of autonomy in which the reasonable frames the rational. Though we cannot look at it at this point, the ground has

⁵⁷ TJ 12.

⁵⁸ TJ 13

already been prepared for the much larger argument, which is that rational choice can provide a foundation for the moral law.

I think it is also of some interest that Rawls interprets the second formulation of the categorical imperative in this way: that is, as the notion that we must heed and respect humanity in ourselves and others because in doing so we respect our character as reasonable and rational beings. Thus our free nature, interpreted as rational, is always in some sense subjugated by our equal nature, interpreted as reasonable. This is the meaning of the second formulation of the categorical imperative as respect for persons.⁵⁹

We should note at this point that in the original position we are still dealing with a hypothetical 'person'. Rawls no less than Kant believes that the agents who perform the bargaining or the reflection about the categorical imperative are ideal.⁶⁰ Thus the reality of the two principles of justice is still in doubt, just as is the reality of the rational agent who employs the categorical imperative in Kant. All Rawls is saying at this point is that, in order for rational deliberation to take place we must presuppose a free and equal agent.

The Reasonable and the Rational

Rawls characterizes his project in "Kantian Constructivism in Moral Theory" as the attempt to: "establish a suitable connection between a particular

⁵⁹ Rawls. "Themes in Kant's Moral Philosophy". 505.

conception of the person and the first principle of justice, by means of the procedure of construction”.⁶¹ This means that Rawls attempts to construct a philosophically coherent story about how the idealized conception of the person as reasonable and rational, can lead to a set of public institutions of justice we all can endorse. Before we examine what Rawls means by constructivism, we must understand what he means more exactly by the reasonable and the rational.

In political terms this means:

[W]henver a sufficient basis for agreement among citizens is not presently known, or recognized, the task to justify a conception of justice becomes: how can people settle on a conception of justice, to serve this social role [of admissible social institutions], that is (most) reasonable for them in virtue of how they conceive of their persons and construe the general features of social cooperation among persons so regarded? (KC 305)

Rawls has in mind here that people reflecting on the issue of justice will see themselves as free and rational and see others in this way as well. Their reflection will be something like the model of reason employed in game theory (where each player tries to maximize his or her advantage but realizes that this is only possible when the other players are treated fairly). In order for such rational deliberation to be effective at developing rules all can agree to, the rational actors will be forced to abandon their personal preferences and incorporate only those rules which apply to all equally. This will make the system morally neutral, though not amoral.

⁶⁰ Kant. GMS 407. O'Neill. "The Method of *A Theory of Justice*". 33-34.

⁶¹ Rawls. "Kantian Constructivism in Moral Theory". 304. Henceforth KC, when cited in the text or "Kantian Constructivism" when cited in footnotes.

To put the issue slightly differently, we could say that the hypothetical imperatives each person at the bargaining table wishes to realize are limited by the recognition that each of the bargainers is equal and that it is thus unreasonable for one member to insist that the group agree to make an exception for that member. Thus the reasonable which models the demands of universality in the categorical imperative frames the debate about which particular hypothetical imperatives can be realized. The notion of universality, which Rawls interprets as equality, frames and restricts the particular rational plan of any actor. This turns classical liberal 'negative' freedom into a more communal 'positive' freedom. Thus when Rawls says that the original position is morally neutral, he means that there is no conception of the good involved in decision making itself. Morality, however, *is* in play in the sense that freedom and equality have a particular moral perspective, which is that the reasonable frames the rational.

But Rawls has not, as yet, given an argument for this overlaying of the rational by the reasonable. In a sense we are asking how equality can come to dominate freedom, that is, to anticipate a bit, what makes up Rawls' argument to the effect that Kant's doctrine of respect for persons will trump a libertarian notion of freedom even in the seemingly libertarian theory of rational choice. This is a problem for Rawls since, taken on its own, as Rawls readily admits, rational choice might, depending on the moral character of its actors, yield a Hobbesian state as readily as the Kantian Kingdom of Ends. There is nothing keeping the rational actors from agreeing to the rules only for prudential reasons and

upholding them only until the relations of power shift in their favor. Rational choice is a method of coordinating people, but it cannot guarantee that people will forever be happy with the position they have been given, even if they initially agreed to it.⁶²

In order to provide stability for the rules agreed to in the original position, Rawls must now reintroduce something like Kant's belief that rationality includes respect for persons and not just the maximization of one's advantage (material or otherwise). While Kant believes rationality and social responsibility to be of a piece, Rawls has here, in the interest of overcoming some of Kant's metaphysical presuppositions, divided them. Rawls has based the laws of his society on issue of rational self-interested coordination among its members. Now he must show that rational self-interest is also in the interest of others and is, in fact, worth agreeing to for 'the right reasons' as he will say in *Political Liberalism*.

Agreement for the 'right reasons' is central to Rawls' overall project since it is meant to show that there actually is real or, one might say, metaphysically true, agreement between different points of view. Of course, Rawls wants to do without the language of truth and is thus forced to explain 'right reasons' through the model of conversation. The problem remains a real one, however, since for there to be agreement, there must be a fundamental intersubjectivity which Rawls just assumes throughout.

⁶² This is a concern not adequately dealt with in *A Theory of Justice*. Other issues, like holding out for a larger share of the pot, have already been addressed. See chapter 3 in *A Theory of Justice*.

In order for this rationality to yield more than prudential agreement or a *modus vivendi*, Rawls must show that having a thin theory of the good allows each agent to move to a thick theory of the good. This is the point of introducing the distinction between the rational and the reasonable.⁶³ Rawls wants to show that instrumental reason as employed in the original position can be seen as an ethical capacity from a different perspective. This leads to a reinterpretation of the original position in “Kantian Constructivism”, which relies more heavily on the notion of equality than its predecessor in *A Theory of Justice* did.

The movement occurs in three stages. It starts from rational autonomy (bargaining proper), moving to full autonomy (bargaining with reasonable or moral constraints) and finally ending up with the readers of Rawls’ theory themselves (which finds its justification in the reflective equilibrium). What Rawls calls the rational or rational autonomy is modeled in pure procedural justice. At the second stage, of full autonomy, Rawls adds to the conception of the person as free and equal two moral powers and two higher-order interests. The first power is that of having an effective sense of justice, the second is the power to form and revise and rationally pursue a conception of the good. Corresponding to these are the higher-order interests of realizing and exercising these powers. (KC 312)

The move to full autonomy and the reasonable, Rawls writes, is “expressed by the framework of constraints within which the deliberations of the

⁶³ See Rawls “Themes in Kantian Moral Philosophy”. 503-504, and “Kantian Constructivism” 316. Also Baynes. *The Normative Grounds of Social Criticism*. 57.

parties (as rationally autonomous agents of construction) takes place”. (KC 317)

This framework is the reasonable ideal of fair cooperation. The framework, by which Rawls means the addition of the two moral conceptions of the person, reciprocity and mutuality, ensure that the plan of the good each person articulates for him or herself also includes the good of others. This is the doctrine of respect for persons as it is expressed in Kant’s second formulation of the categorical imperative, the formula of humanity as an end in itself. Thus the two moral powers overlay the process of rational deliberation, transforming the instrumental deliberative process in the original position into a process of mutual recognition and fair cooperation. Rawls elaborates: “In justice as fairness, the Reasonable frames the Rational and is derived from a conception of moral persons as free and equal. Once this is understood, the constraints of the original position are no longer external.” (KC 319) I take this to mean that only the device of the original position (which models instrumental reason) imposes the constraint of fair cooperation on the people. For the people in the original position, social cooperation is not intuitive. But it is so for fully autonomous people who live in the institutions which the two principles of justice have helped to create. For they see themselves as possessing the two moral powers and thus restrict their pursuit of the good in the name of something more than the maximization of their material gain.

The movement of the two stages so far trades on the distinction between different perspectives. If we move back a little, we might recall that the purpose of the original position is to develop principles of justice out of our

presuppositions about moral character. That is, what kind of laws would free and equal people come up with if left to their own devices? What Rawls does is to draw out first what free or rational individuals would do and then to overlay this with what people who are both free and reasonable would do. Rational people seek to maximize their benefit while reasonable people seek to maximize their benefit with the concerns of others in mind. This parallels exactly the structure that Kant argues for as well: we are rational beings insofar as we try to realize our ends by adopting the means to do so, but we are moral insofar as we adopt only those ends which we can will others to adopt as well.

Thus Rawls can say: “The unity of practical reason is expressed by defining the Reasonable to frame the Rational and to subordinate it absolutely; that is, the principles of justice that are agreed to are lexically prior to their application in a well-ordered society to claims of the good.” (KC 319)

The lexical ordering of the reasonable over the rational also parallels Kant’s division of practical reason into empirical practical reason and pure practical reason. While empirical practical reason— the hypothetical imperative— means acting according to any practical principle, pure practical reason— the categorical imperative— means acting according to the principle of the moral law.

This parallel notwithstanding, it is still unclear whether our wills can actually act according to the moral law. That is, while it may be agreed that it is analytic that we are rational, it has not yet been shown that we are reasonable. Furthermore, we now face the problem, rightly noted by Rawls, of whether what appears to us as descriptively reasonable (i.e. apparent respect for persons) is

actually respect for persons or merely arises for prudential reasons. (KC 321) For Kant the answer to the lexical ordering of the reasonable over the rational comes from the argument about the priority of autonomy over heteronomy. And the same goes for Rawls. Autonomy simply means following your own reason and this is what we do when we reason from laws which we ourselves have given ourselves.

This point also brings out a little bit of the ambiguity in 'rational choice' which requires it to be subjugated under the 'reasonable'. Just as a maxim or plan of action might or might not originate from pure practical reason, so too the results arrived at in the original position might or might not be expressions of a universally applicable law. They might just be agreed to because it is easiest for all. What concerns Kant and Rawls clearly is that the practical principle that is adopted be adopted because it is universal and thus for the sake of reason itself, rather than for expedient reasons. Thus, in order for us to be sure reasons are adopted for the sake of reason, we must put constraints on our adoption of maxims. We should adopt only those principles which are free of contingent concerns.

Before moving on, let us recap the argument so far. We started with the notion of the original position as an interpretation of the categorical imperative. This was meant to show that the bargaining in the original position together with the veil of ignorance models Kant's twin notions of the moral law and autonomy. These arguments were essentially meant to show that if we are autonomous, we will choose ends that are universally acceptable.

The second step was for Rawls to draw out the presuppositions behind the notion of the *vernünftig* person in Kant who is able to reflect and act in accordance with the categorical imperative. This meant seeing the double structure of the agent's practical reason, which revealed itself to be both pure and empirical. Rawls characterized the agent as 'reasonable and rational', where the reasonable (modeling the categorical imperative) absolutely subjugates the rational (modeling the hypothetical imperative). This conception of personhood was then linked to Rawls' conception of the person as free and equal.

There are still two elements missing from this argument. The first, to which we will now turn, is the question of how we get from the presupposed character of the agent as reasonable and rational to the content of the principle of justice, which so far has been described only formally. The second question, which we will come to after that, is what justifies the assumption of people as 'reasonable and rational' in the sense of being free to set their own goals.

Constructivism

Constructivism is meant to be the way to get from a certain conception of the person (here, free and equal) to the appropriate principles of action for such a person. This means that constructivism seeks to draw out the content of the conception of the agent and to formalize it. That is, if the CI-procedure is the appropriate form of a rational principle, what is the appropriate material? The

answer is the free and equal agent.⁶⁴ It is the answer to the question: what should we do when we act under the moral law or use our pure practical reason (which amounts to the same thing)?

In Kantian terms this means that: “the totality of particular categorical imperatives [...] that pass the test of the CI-procedure are seen as constructed by a procedure of construction worked through by *rational* agents subject to various *reasonable* constraints.”⁶⁵ Each time we reflect and determine a law for ourselves we construct an element in a universal set of rules which can then be abstracted and turned into a general duty. Rawls’ two principles of justice are a version of what might be arrived at in such an abstraction. The point, though, is that the maxims of conduct permitted or enjoined by rational reflection are not theoretical speculations; they are responses to actual needs for clarification of the permissibility of intended action.

O’Neill notes that the constructivist position is anti-realist because it denies that moral facts are discoverable in theoretical terms.⁶⁶ Constructivists believe that ethical principles are constructed by human agents, that these principles are practical and that they are objective. To this effect O’Neill quotes Rawls as saying in “Kantian Constructivism” that: “Kantian constructivism holds that moral objectivity is to be understood in terms of a suitably constructing social

⁶⁴ For this way of putting the problem see Christine Korsgaard. “Kant’s Formula of Humanity”. 107.

⁶⁵ Rawls. “Themes in Kant’s Moral Philosophy”. 513-14.

⁶⁶ O’Neill. “Construction in Rawls and Kant”. See also Christine Korsgaard. “Reasons We Can Share”.

point of view that all can accept. Apart from the process of constructing the principles of justice, there are no moral facts". (KC 307) This is because moral objectivity stems from moral reflection by individuals. If the maxim is in accord with the moral law then it is permissible or enjoined. But since the moral law legislates universally, the results are valid for all and thus objectively true for all rational and reasonable agents. The only way objectivity can be understood is by referring to the set of agents who fall under it. There are no independent moral facts because they have no way of reaching us, just like the thing in itself.

This contrasts notably with the moral realist position in which principles are regarded as discoverable and the truth of a subjective moral claim rests on the claim of being in accordance with an objectively true moral value. Rawls thus claims that Kant regarded rational intuitionism, for example Leibniz, as heteronomous because it held that principles of morality are true in an objective sense. This is heteronomy because, if this were true, we would have to subject ourselves to a law which is external. Thus Rawls claims that:

in Kant's moral constructivism it suffices for heteronomy that first principles obtain in virtue of relations among objects the nature of which is not affected or determined by our conception of ourselves as reasonable and rational persons (as possessors of the power of practical reason), and of the public role of moral principles in a society of such persons. [...] Kant's idea of autonomy requires that there exists no moral order prior to and independent of those conceptions that is to determine the form of the procedure that specifies the content of first principles of right and justice among free and equal persons.

And a little later:

"[This thesis] simply means that the form and structure of this procedure express the requirements of practical reason. These requirements are embedded in our

conception of persons as reasonable and rational, and as the basic units of agency and responsibility.”⁶⁷

To recap then, we can say that constructivism is the idea that the content of our highest moral principles stems from the rational and reasonable reflection upon our concepts as free and equal agents. Constructivism models autonomy in the sense that it constitutes the moral law or principle of justice from within its own rational and reasonable reflection. Nothing can count as a law for me without my having determined it for myself. This strongly echoes Kant’s claim that there is nothing good in itself except the good will.

Now, as before, there is here an emphasis on the first person perspective. That is, constructivism is just the CI-procedure insofar as it pertains to determining the content of the moral law. The content of the moral law has the content it has because I have (rationally) reflected upon it and have determined that it has this content. We must, however, keep open the possibility that when this first person perspective is switched to a third person perspective, as it is in rational choice, we lose normativity altogether. We will return to this issue.

An important consequence of the constructivist method is that the terms of justice are procedural, that is, no contentful definition of justice can ever be given. There are two types of procedural justice: perfect procedural justice where a good is in sight and we work toward achieving or actualizing it, and pure procedural justice. Rawls writes of the latter: “The essential feature of pure procedural justice, as opposed to perfect procedural justice, is that there exists

⁶⁷ Rawls. “Themes in Kant’s Moral Philosophy”. 512. Henceforth ‘Themes’ when cited in the text.

no independent criterion of justice; what is just is defined by the outcome of the process itself.” (KC 311)⁶⁸ The point Rawls wishes to emphasize here is that: “there exists no standpoint external to the parties’ own perspective from which they are constrained by prior and independent principles in questions of justice that arise among them as members of one society”. (KC 311)

The contrast again is with the realist or rational intuitionist position, in which knowledge of the good is applied to a certain situation to bring the good about. Pure proceduralism, by contrast, is practical and constitutes the good as it develops. Or, as Rawls writes in *A Theory of Justice*, “A procedure translates its fairness to the outcome only when it is actually carried out.”⁶⁹ This is a uniquely practical conception of justice which has no theoretical analogue.

In a sense, one would be hard pressed to say why something is just other than that it has been rationally and reasonably agreed to. The constructivist procedure is thus completely immanent. Moral objectivity is reached by a process of internal reflection in which the maxim is referred to an internal concept for adjudication and is implemented on the authority of the process of deliberation alone. Objectivity is not sought in external metaphysical truths but arrived at through thinking itself. So for constructivism, truth is constructed in rational engagement with the world rather than found by poking around in the world.

⁶⁸ See also *A Theory of Justice* where Rawls writes: “[P]ure procedural justice obtains when there is no independent criterion for the right result: instead there is a correct or fair procedure such that the outcome is likewise correct or fair, whatever it is, provided that the procedure has been properly followed.” (TJ 75)

⁶⁹ Rawls. *A Theory of Justice*. 75.

Construction thus has two elements. First, it is a process internal to the agent and as such it is from a first person perspective. No one can reflect for me. Second, it is practical. Since reflection on the permissibility of performing an action stems from an incentive for action, the result of my reflection can only ever be manifested in my action itself. The result of my reflection can only ever be what I actually do, that is, what motivates me. If I say I ought to give \$100 to charity and do not, I have actually decided to keep the \$100. A practical constructivism thus relies on a notion of pure practical reason, that is, the idea that we are capable of reflecting on our ends by the use of the moral law or the two principles of justice.

Let us take a step back again and see where the argument has gotten us so far. Constructivism and proceduralism were introduced to provide a link between the conception of persons as free and equal, or as reasonable and rational and the content of the principles of justice. Rawls' contention was that through the process of construction, or through autonomous reflection, these ideal agents would determine a set of principles which are able to govern the agents who have developed them in fair cooperation. Construction was then the way to bring out the content of the basic idea of the reasonable and rational agent without introducing any alien conceptions of how the world is or ought to be. The only tool available to the reasonable and rational agent in determining what the principles of justice are is reason. We also found that this constructivism proceeds from a practical point of view, which cannot be justified in theoretical

terms. Justice is immanently constituted as doing that to which all involved have agreed.

Thus three of the four elements of Rawls' argument are in place. The original position has been established as yielding a universal principle. The presuppositions about the moral character (freedom and equality, reasonable and rational) of the agents who participate in the original position have been examined. And lastly, constructivism has presented a way for us to move from these presuppositions of moral character to the actual content of the formal characteristics of the moral law: the principles of justice.

The only element that is still missing is the justification of why we should think that we are actually those people in the original position who frame the rational by the reasonable.

Excursus: Rawls and Kant, parallel arguments

I have argued that Rawls has roughly followed the structure of the *Groundwork*. He has developed the categorical imperative in terms of the universality of the two principles of justice, corresponding to the first formulation of the categorical imperative. Then Rawls has switched perspectives and has argued that the presupposition for such a law is that people respect their humanity and, finally, Rawls has contended that in order to act under the moral law, we must imagine ourselves as instantiating the two principles of justice. For Rawls this is the kingdom of ends. Let us examine these parallels in a little more detail.

Rawls writes: “In the first formulation [of the categorical imperative], which is the strict method, we look at our maxim from our point of view. [...] We are to regard ourselves as subject to the moral law and we want to know what it requires of us.” (Themes 505) This, I want to argue, is similar to the original position in which we want to know the formal structure a principle of justice would have.

In the second formulation, however, we are to consider our maxim from the point of view of our humanity as the fundamental element in our person demanding our respect, or from the point of view of other persons who will be affected by our actions. Humanity both in ourselves and in others is regarded as *passive*: as that which will be affected by what we do. (Themes 505)

As I have already indicated above, I take this to be the perspective of drawing out the presuppositions about agents in the original position. To frame the rational by the reasonable means to see ourselves as passive in the face of the hypothetical imperative and to try to avoid damage to our humanity by restricting its scope. Our humanity is the material for the application of the CI-procedure in the sense that this is the purpose for that procedure. Rawls adds: “The point is simply that all persons affected [by my will] must apply [the CI-procedure] in the same way both to accept and to reject the same maxims. This ensures a universal agreement which prepares the way for the third formulation.” (Themes 505)

“In [the third] formulation we come back again to the agent’s point of view, but this time we no longer regard ourselves as someone who is subject to the

moral law but as someone who makes the law. The CI-procedure is seen as the procedure adherence to which, with a full grasp of its meaning, enables us to regard ourselves as legislators— as those who make universal public law for a possible moral community.” (Themes 506) This last formulation is clearly analogous to constructivism in the sense that in constructivism we develop positive law out of our conception of ourselves as free and equal.

I provide this juxtaposition of the structure of Rawls’ and Kant’s arguments not only to support Rawls’ claim that *A Theory of Justice* is largely Kantian in orientation but to show that *A Theory of Justice* brings out central features of constructivism which must be seen as not just incidental but substantive contributions to Kant scholarship.⁷⁰ I further wish to argue that by tying his theory to Kant so closely, Rawls’ theory is subject to many of the same difficulties as Kant’s work. These difficulties have mainly to do with the problem of justification. For instance, the failure of Kant’s deduction of morality has left Kant without a footing from which to say that humans are indeed able to interact respectfully with one another. The two commentaries on Kant which we looked at in the previous chapter attempted to push Kant toward a more historical and thus immanent view of the interaction between agents and nature. Because Rawls avoids this push toward immanence and stays at what might be called the ‘common sense’ level, he also lacks a philosophically rigorous conception of intersubjectivity. Rawls’ rejection of metaphysics, as I have said before, leaves

⁷⁰ See Rawls. *A Theory of Justice*. §40.

him without an answer to the question of how people can actually be relied upon to treat each other with respect.

Justification and the Reflective Equilibrium

If the theory of construction is the justification for the two principles of justice, then what justifies construction? Rawls' answer, like Kant's answer to the problem of why humans should consider themselves free, is quite simply that constructivism is not justified in a theoretical way, but is given its authentication through cohesion into the perspective of existing humans who find that they agree with it. Justification is given through action. This notion of coherence is the final step in the three part development of authentication presented in "Kantian Constructivism".

We have already seen the first two steps. The first was that of the actors in the original position. They were rationally autonomous in the sense that their understanding of their life and the good was hidden behind the veil of ignorance. Then came the second perspective, that of full autonomy in which the perspective was that of the agents stepping back from rational autonomy in order to implement the second level of insight, reasonableness or the sense of justice. This introduced the categorical imperative to frame the hypothetical imperative, but the agents who employed it were still only ideal. What is needed, as I have said, is to determine whether the ideal agents of full autonomy bear any resemblance to us.

Finally, Rawls comes to consider the last perspective, “that of ourselves—you and me—who are examining justice as fairness as a basis for a conception of justice that may yield a suitable understanding of freedom and equality” (for our own practical use). (KC 320-21) Rawls continues: Here [in the third perspective] the test is that of general and wide reflective equilibrium, that is, how well the view as a whole meshes with and articulates our more firm considered convictions. [...] A doctrine that meets this criterion is the doctrine that, so far as we can now ascertain, is the most reasonable for us. (KC 321)

At this third perspective then, we have arrived at the criterion for a final justification of Rawls’ theory. The problem for Rawls, as for Kant, is that we cannot prove that people believe themselves to be those ideal agents. Rawls is quite convinced that the failure of Kant’s deduction of the moral law is sufficient to show that such an idealized theory approach makes no sense. So, according to the third perspective, the justification or authentication comes down to what Rawls calls the reflective equilibrium.

Let us now examine what the reflective equilibrium is in more detail. As Kenneth Baynes puts it: “reflective equilibrium refers to a condition in which an individual’s concrete moral judgments have been brought into harmony with her higher-order moral principles”.⁷¹ This harmonization occurs first through a narrow process of reflective equilibrium in which one moves back and forth between concrete judgments (in, say, the manner of the categorical imperative in which

the subject decides on a maxim and, using the categorical imperative procedure, determines whether it can be acted upon— if not, a new maxim must be created and tested) and then through the wide process of reflective equilibrium in which one's own judgments are bought into harmony with general social norms, shared by most readers of *A Theory of Justice*.⁷²

Let us now turn to Rawls' own characterization of the process before turning to criticisms and defenses of this method. Rawls holds that his theory of justice describes our own sense of justice. (TJ 41) The justifications of his theory of justice, modeled by the original position and background conditions, are all reflections of our own considered judgments. The need to write *A Theory of Justice* in the first place, however, must have been generated by the knowledge that, on the face of it, not everyone currently *does* in fact share Rawls' conception of justice. The task of justifying the theory of justice thus must occur through a process of fleshing out those beliefs we actually all hold.⁷³

The process of achieving reflective equilibrium systematizes our beliefs.⁷⁴ What we do in the narrow reflective equilibrium is thus similar to what we do in

⁷¹ Baynes. *The Normative Grounds of Social Criticism*. 69.

⁷² Rawls already characterizes the discussion about Justice as Fairness as taking place within a bounded society, one that endorses liberal democracy. Readers coming from outside this realm, may not agree with him.

⁷³ Who exactly the 'we' is, has been the subject of much debate. See, for instance, Susan Moller Okin. "*Political Liberalism, Justice, and Gender*".

⁷⁴ See Rawls' formulation about the original position: "The conditions embodied in the description of the original position are ones that we do in fact accept. Or, if we do not, then perhaps we can be persuaded to do so by philosophical reflection." (TJ 19)

the CI-procedure. We take a practical problem which, admittedly, is more abstract than our everyday practical concerns, and reflect on it. For what is systematizing but bringing disparate concepts under a general principle of practical reason.

Rawls is thus arguing that through the reflective equilibrium we might see our own principles of reason to be the same ones that Rawls, however, has advocated in *A Theory of Justice*. This will only produce coherence and not more because it cannot be argued (made explicit theoretically) that we *must* share the same principle of pure reason. *A Theory of Justice* is thus an invitation to reflect on a deeper problem of our time and to take Rawls' theory as an occasion for reflection.

There is thus a positive and a normative side to the reflective equilibrium, or, as Scanlon put it, a descriptive and a deliberative side. As a method of arriving at an accurate portrait of justice, we must dig within ourselves to find normative notions we endorse.⁷⁵ Both sides seem to be included in the following statement by Rawls: "we do not understand our sense of justice until we know in some systematic way covering a wide range of cases what these principles are." (TJ 41) Indeed, in this statement of the purpose of the reflective equilibrium it is not possible to separate the two senses. Since, however, the process of the reflective equilibrium is a theoretical undertaking to which we subject our considered judgments, it seems appropriate to call it a method of deliberation.

⁷⁵ Scanlon. "Rawls on Justification". 145.

The method itself is not explained in great detail in *A Theory of Justice*.⁷⁶ I will thus cite only the two main passages from this work in which Rawls describes the reflective equilibrium process as it pertains to original position:

By going back and forth, sometimes altering the condition of the contractual circumstances [in the original position], at others withdrawing our judgments and conforming them to principles, I assume that eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted. (TJ 18)

A conception of justice cannot be deduced from self-evident premises or conditions on principle; instead, its justification is a method of the mutual support of many considerations, of everything fitting together into one coherent view. (TJ 19)

The reflective equilibrium begins with our considered opinions which must be made “under conditions favorable to deliberation and judgment in general”. (TJ 40) According to Scanlon considered moral judgments are of three types: (1) Those that have to do with how the judgment is made (knowledge of the issues, the person neither gains nor loses by the outcome), (2) those that have to do with

⁷⁶ Rawls’ idea of the reflective equilibrium has been taken up in the fields of moral philosophy and the philosophy of science. See, for instance, the more rigorous formulation James Blanchowicz gives (which is not based strictly on Rawls’ account). Likening the reflective equilibrium process to repairing a ship at sea, Blanchowicz writes: “It is not just the fact that one is resting on a dry part of the ship in one’s efforts to repair a leaking part and that one may later rest on the repaired (formerly leaking) part to repair a new leaking (formerly dry) part that establishes genuine reflective equilibrium, but rather the fact that the way in which one rests on these respective parts

how the judgment is held (the level of confidence in the judgment and how stable the judgment is over time), and (3) those which should be intuitive with respect to ethical principles.⁷⁷ A central feature of these conditions, and one which will be relevant to criticism of the reflective equilibrium discussed below, is that the process of the reflective equilibrium is always subject-dependent. That is, it is always my judgment that comes in for consideration. In this sense, judgments are judgments only when they come with my reasons for the judgments attached.⁷⁸ They are thus not comparable to observational data.⁷⁹

As a coherentist strategy, convergence in reflective equilibrium is only evidence of how much agreement we already have. It is not normative, in the sense that it might convince one who does not hold what I hold to change his or her mind. It is purely introspective. This is because, as Norman Daniels argues, coherentism in the form of the reflective equilibrium remains agnostic about whether there is any truth which it might approximate. This agrees with the point about anti-realism raised earlier according to which constructivism develops all 'truths' through reflection itself.

is different in each case[...]" Blanchowicz. "Reciprocal Justification in Science and Moral Theory". 460.

⁷⁷ Scanlon. "Rawls on Justification". 143.

⁷⁸ This is perhaps the place to note that Rawls never develops an adequate justification of the reflective equilibrium from the first person perspective, and thus ultimately leaves himself open to criticism from Kantians and others who regard the subject as the primary unity of ethical coherence. See the discussion of Christine Korsgaard and Onora O'Neill below. Both seek to remedy this deficiency in Rawls' account through their respective theories of practical reason.

Problems with the Reflective Equilibrium

As we saw above, the method of the reflective equilibrium is a way of becoming clear about one's own ethical convictions. We examine our thoughts and our principles are measured against what we might have read or discussed with others. The important point to keep in mind is that we are now ourselves, comprehensive subjects with commitments to notions of the good. This means that each of us reflects from a different perspective.⁸⁰

At this point we must, however, make a distinction which I mentioned earlier, namely the distinction between the content of the theory and its very possibility. For there is an ambiguity in the charge that we reflect from different perspectives as real existing agents. The charge might mean that, since we are different, we are not sure whether we will come to the same conclusions as Rawls does. But it might also mean that we would have completely different conceptions of morality or that morality might be denied altogether. The former

⁷⁹ See Norman Daniels. "Wide Reflective Equilibrium and Theory Acceptance in Ethics". pp. 169-72. See also Brandt. "Wide Reflective Equilibrium".

⁸⁰ This is what Sandel has in mind when he argues that the Rawlsian deontological subject is incapable of normative commitments because he or she has been cleansed of all contingency which would necessitate normativity in the form of judgment. In order to make the Rawlsian subject capable of normativity, normativity must be introduced at a later stage but this is impossible given the thinness of the subject as it is conceived in the original position. Sandel. *Liberalism and the Limits of Justice*.

point is addressed by the bulk of Rawls' argument while the latter point refers to a problem Rawls does not have much to say about.⁸¹

The problem of different starting points for reflection brings with it a host of problems for Rawls. For instance, there is no longer any compelling connection between the perspectives of the different subjects being asked to endorse Justice as Fairness. The point is put nicely by Baynes who argues that there seems to be no reason for me to accept the results of the reflective equilibrium unless I am the one who has undergone the process myself. This, presumably, is what Rawls means when he writes that "each person has in himself the whole form of a moral conception". (TJ 44) Here, whole must mean complete for me and not, as in Kant, universal. Thus, there does not seem to be any reason why I should be swayed by a subjective process of reasoning not my own.⁸² We are thus back at the question of who the 'we' who endorses the considered moral judgments is and whether there is any connection among the individuals which make up the 'we'. In *A Theory of Justice* and in "Kantian Constructivism", this 'we' seems not to have been theorized at all where possible justification is concerned.

Scanlon, similar to Baynes, argues that the reflective equilibrium process is normatively underdetermined. This charge states simply that no conclusive evidence for or against Rawls' theory can be gotten from a coherentist

⁸¹ Rawls addresses this issue in *Political Liberalism* where he talks about commitment to the liberal state as opposed to the *modus vivendi*, a temporary commitment which, in certain extreme cases, might seek to overthrow the whole system.

justification.⁸³ Since the reflective equilibrium does not offer a determinate process by which one might arrive at ethical conclusions, it is quite possible for two people to start from the same premises and, using the reflective equilibrium method, still arrive at different conclusions. Rawls acknowledges this point when he says that his theory of justice is just 'a' theory of justice.⁸⁴ (TJ 43-44) But as a theory of justice it must include the claim that something is normative for us even if we cannot agree entirely on what it is. The problem is thus that a coherentist theory which seeks its justification in the reflective equilibrium is too weak to bind people of differing perspectives together because it cannot on its own overcome the differences that people with previous normative commitments bring to bear on their reflections. Coherentism, in other words, seems not to be able to provide consensus where there is none to begin with.

There is another, deeper objection here, however. Scanlon has argued that someone's employment of the reflective equilibrium commits the evaluator of

⁸² See Baynes. *The Normative Grounds of Social Criticism*. 74.

⁸³ Brandt argues, for instance, that Rawls' argument comes to a conclusion no more forceful than that: "A coherent set of beliefs can be made more convincing than another set even if there is nothing which can confirm or refute it." "The Science of Man and Wide Reflective Equilibrium". 272-73.

⁸⁴ Concerning the intersubjectivity of the reflective equilibrium process, Rawls writes that the question must remain open: "I shall not even ask whether the principles that characterize one person's considered judgments are the same as those that characterize another's. I shall take for granted that these principles are either approximately the same for persons whose judgments are in reflective equilibrium, or if not, that their judgments divide a few main lines represented by the family of traditional [moral] doctrines [...]". (TJ 44) Rawls adds, referring to himself, that: "if we

the argument who undertakes it to nothing at all.⁸⁵ This question delves deeper since it asks the more fundamental question of whether morality exists at all and thus lays bare the assumption Rawls has so far been making about the reflective equilibrium, namely that it *is* the pure employment of practical reason. If the reflective equilibrium is, in fact, the pure employment of practical reason, there will be no problem with coherence beyond the merely technical problem of the correct assessment of the facts. We thus need some further argument about why the reflective equilibrium is, in fact, the employment of pure practical reason and not some other principle. This goes to the more fundamental question of the possibility of morality and thus relates quite clearly to Kant's own failed attempt at proving intersubjectivity.

The problem I am here insisting on is that that the answer to the problem of the justification of the two principles of justice in the reflective equilibrium cannot be gotten through an analysis of the coherence of the two principles of justice with our own perspective as readers of political theory through the reflective equilibrium. The deeper problem suggested here turns on the question about the possibility of morality in general, which cannot be answered by coherentism precisely because it is a question of first principles or metaphysics, if you will. Indeed, coherentism can only give an evaluation of the rightness or

can characterize one (educated) person's sense of justice, we might have a good beginning toward a theory of justice". (TJ 44)

⁸⁵ See Scanlon. "Rawls on Justification". 152 and O'Neill. "Constructivism in Ethics". 206-207.

justice of two principles of justice if it is assumed that coherence is really an expression of morality or practical reason.

Before taking up this final issue, we must look a little more closely at what the role of pure practical reason is in Rawls' theory. And this crucially depends on the perspective employed in the philosophical reasoning of *A Theory of Justice*.

Pure Practical Reason and the First Person Perspective

We have now seen all four elements of Rawls' theory so I now want to take stock of the argument as a whole and make good on the promises for elaboration I made during the reconstruction of the argument. I will thus discuss what I see as the real problem in Rawls' ultimate justification of his theory, by which I mean the position of pure practical reason. In both the original position and the reflective equilibrium Rawls presents us with a conception of normativity, through bargaining and the interpretation of social norms, which seems to want to sidestep the question of the need for a justification of his claim for our ability to employ pure practical reason. I will argue, however, that a notion of pure practical reason must underlie both conceptions. I will then return to the issue of whether pure practical reason receives a foundation in Rawls' work.

The first problem I mentioned was the problem of what I argued was the substitution of the original position for the categorical imperative in *A Theory of Justice*. I noted that in this move Rawls replaced a first person perspective with a third person perspective. He seemed to be arguing that the process of deliberation under the veil of ignorance was just as good at leading to the two

principles of justice as solitary reflection. Indeed, the substitution rather suggests that Rawls thinks rational choice is a better model for ethical thought than solitary reflection.

From a Kantian perspective, however, this move seems highly suspect. For what gives rise to normativity in Kant is that I make the law for myself, that I am an autonomous actor. As is clear from the reading I gave above, Rawls also considers this to be the case with the agents in the original position behind the veil of ignorance. But rational reflection, as Kant sees it, operates only from the first person perspective. That is, something is normative for me because I choose to adopt it as a principle. No one else can make me adopt as my end something I do not freely choose as an end. You may force me to do it, but it will not be my end.

This is just the familiar point that practical reason cannot be given a theoretical explanation. No one can convince me by argument that I should adopt their reasons. I must convince myself. So, if deliberation in the original position is really a 'compromise' as Rawls states, then the agreement reached in it is not normative for anyone since it does not represent a principle anyone actually endorses. The principle that has arisen through the compromise might, of course, still be adopted, but Rawls has not given us any argument for why those in the original position should adopt the principles they have reached in negotiation.⁸⁶

There is a way out of this argument, of course. It is essentially that the original position with its multiple parties is just a way of representing what goes

on in rational reflection in the CI-procedure. The move from the CI-procedure to the original position is just heuristic.⁸⁷ That this is so becomes quite clear, I think, when one examines the notion of constructivism, which is meant to connect the two principles of justice to the original position. Constructivism seeks to draw out the consequences of our presuppositions about the agents negotiating in the original position. But in order for us to be able to draw out anything about them, we must assume that they have something in common, namely the concepts of freedom and equality. This is why Rawls refers to these agents as idealized. In order for the process of construction to yield anything at all, 'idealized' must mean that they are at least generally the same. If this is so, then the move from free and equal individuals through construction to the two principles of justice merely mirrors Kant's movement from the *vernünftig* individual through rational reflection to the moral law.

As such, it is no mystery that the agents in the original position can come to a 'compromise' which is normative for all. The compromise is no compromise, it is really the presupposition of the moral theory underlying the make up of the agents— justice as fairness. There has thus been no shift from the first person perspective which admits of the use of practical reason to the third person

⁸⁶ Rawls. *A theory of Justice*. 104-105.

⁸⁷ See Ronald Dworkin. "The Original Position". Dworkin essentially argues that the original position is just a way of testing our intuitions about justice against a general theory. As such, of course, it bears close resemblance to what Kant had in mind with the CI-procedure. Rawls admits as much himself when he states that collective bargaining is really just a representation of the process of reflection employed by the noumenal self. *A Theory of Justice*. 226.

perspective. There is also thus no issue of convincing anyone of the rightness of the two principles of justice.

So, as I think I have shown, the problem of normativity of the two principles of justice does not arise at the level of the original position since, fundamentally, the original position models the use of pure practical reason by an autonomous self. This does not mean, however, that the problem of normativity has been laid to rest. The normativity of the two principles of justice is simply moved back to the reflective equilibrium and to the question of the acceptance of it by comprehensive subjects. It also does not mean that the problem of the first person to third person switch and the problems this entails has gone away.

We saw above that the reflective equilibrium was designed to bring our own considered opinions into accord with those of others. But the reflective equilibrium is meant not only to help us clarify our own views to ourselves, but also to bring these views into harmony with those of others. Rawls, as has been noted, calls these the narrow and wide reflective equilibrium respectively. But while the narrow reflective equilibrium is a process of internal reflection which seeks to help us get clear about our own principles, the wide reflective equilibrium is constructed as a dialogue between different members of society, perhaps modeled on Kant's notion of public reason. There is no problem with this picture unless one regards the two processes of the reflective equilibrium as substantially different.

For, while it is fairly clear, given the possibility of introspection Rawls allows for, that we would become more clear about our principles in the narrow reflective equilibrium, it is more difficult to see how the process of the wide reflective equilibrium would help us to clarify our ideas about morality if we think this process is significantly different from the process of narrow reflective equilibrium. Indeed, on the reading of practical reason which I have been giving, and which I have argued Rawls' reading of Kant endorses, there can be no distinction between these two types of reflection since social norms and reasons could only ever function as candidate reasons which we would then have to endorse ourselves anyway. The entire weight of social convention must still be rationally examined by us if we are to act on its directive while maintaining our autonomy.

According to the arguments I have just given, we must conclude that Rawls' attempt to build greater stability for his system, through both the notion of bargaining in the original position and through the idea of wide reflective equilibrium which ultimately rely on a notion of public reason, is really reducible to the employment of pure practical reason by each individual.

The Problem of Normativity and the Necessity of its Justification

The fact that I have argued that the social anchoring that Rawls wants to give his theory by embedding it in broad social views is inconsistent does not mean, however, that the theory must be rejected or even that its steps are incoherent. I have merely shown that Rawls actually sticks far closer to Kant's

general argumentation than is usually supposed. Two theoretical moves have been rejected but as long as we interpret these moves as merely heuristic, the general theory remains intact. It is thus time to come to the question of the final justification of the reflective equilibrium, in other words, whether there is an account of pure practical reason in Rawls' theory.

And here we come to the central problem of Rawls' justification. Kant saw his theory as hinging on the proof or authentication of the necessity of freedom and morality both in the deduction and in the fact of reason doctrine. Rawls does not think his theory requires such a grounding.

This brings us again to the problem of the first person and third person perspective of practical reason. I argued first that the agents in the original position, as autonomous and idealized, must share the same conception of freedom and equality, and that this means that they are really not substantially distinct in a way that would necessitate a compromise in determining the two principles of justice. Then I argued that the reflective equilibrium process which we must all engage in, in order to determine whether we actually believe ourselves to be similar enough to the idealized agents in the original position to endorse the two principles of justice they determine, also had to stem from a first-person reflection. Thus the claim and its authentication both stem from a first-person perspective.

Without a proof for the necessary identity between the results of the original position and the results of our own reflection, the most that can be said of the two principles of justice is that they cohere. And this is all Rawls wants to say.

Rawls refuses Kant's deduction of morality in favor of Kant's fact of reason. And Rawls interprets the fact of reason as a coherentist justification for the two principles of justice.

Thus Rawls writes:

Pure practical reason is authenticated finally by assuming primacy over speculative reason and by cohering into, and what is more, by *completing* the construction of reason as one unified body of principles: this makes reason self-authenticating as a whole. (Themes 523)⁸⁸

The idea here is that since there can be no theoretical proof of freedom and morality, the only justification for morality that can be given is that we recognize ourselves as moral beings, that is, we recognize ourselves as the agents who participate in the original position. This means, as we saw, that the speculative part of his theory, the original position and the mutual regard of rational autonomy, cannot be justified except through empirical endorsement by fully autonomous actors, the people engaged in ethical reflection— you and me.

O'Neill comments on this passage:

The distinction which Rawls draws here between Kant's constructive use of the [categorical imperative] to establish ethical principles and the merely coherentist justification of the [categorical imperative] itself parallels his own strategy in *A Theory of Justice*, where the principles of justice are constructed using [the original position], but [the original position] itself receives only a coherentist justification.⁸⁹

⁸⁸ See also O'Neill. "Constructivism in Rawls and Kant". pp. 356-57.

⁸⁹ O'Neill. "Constructivism in Rawls and Kant". 357.

By a coherentist justification Rawls means that in recognizing ourselves in the idealized agents of the original position we take their principles as our own principles. The primacy of reason which is invoked in the fact of reason doctrine means, however, that there can never be a theoretical explanation of why we must recognize ourselves in these agents. We just do. And we do so only if we actually act on those two principles of justice, make them ours. For, as the first person perspective requires, it is only through adopting them as principles of action that they can become our principles.

Thus, much rides on the notion of recognizing ourselves in the idealized agents of the original position. This recognition comes down to believing ourselves to be capable of employing pure practical reason. And this belief is what Rawls means by cohering into a much broader notion of reason. Rawls thus offers a minimalist authentication of the possibility of morality itself.

I mentioned earlier that Rawls thought that the categorical imperative offered only modest help in determining contentful principles of justice. We are now in a position to make better sense of this claim. Because of his coherentist justification of the two principles of justice, Rawls does not maintain that *his* two principles of justice are the only ones possible. That is, he does not maintain that he has determined the precise content of the laws our social organization should take. He has proposed 'a theory' of justice which is open to revision.

But this is not a significant departure from Kant, since Kant was not proposing a significant content to the moral law. He was only interested in showing that there is such a thing as the moral law. This, however, is a position

in which Rawls must follow Kant, since in order for there to be any kind of theory of justice, the possibility of a theory of justice must be given. And this is what Kant's deduction and later his doctrine of the fact of reason seeks to show.

The second claim is deeper, for it contains a thesis about the ultimate justification of morality. Here Rawls just assumes that the principle of practical reason really exists. In this sense, the revisability of the two principles of justice which, as we saw, are supposed to be derivatives or incarnations of Kant's categorical imperative, depends on there being such a thing as the categorical imperative or freedom in the first place. By seeking to give a weaker interpretation of the categorical imperative in terms of the CI-procedure, Rawls has given up on Kant's claim that the weak autonomy thesis must be turned into a strong autonomy thesis. Rawls has, in other words, given up on the idea of showing that humans are rational beings and has just assumed that we are.

But giving up on the strong autonomy thesis means that, as I have argued, there is no answer to the question raised in the last chapter, namely, why should we think that we appetitive humans are motivated by rational laws and hence, why should we think that what you think is 'rational' is not just a way of subjugating me. The whole question of justice, in other words, rests on showing that we are all in possession of a common rationality which can help us to overcome our appetitive natures and adhere to derivatives of the categorical imperative as Rawls or anyone else proposes them.

To put the question one last way in terms of Kant's analytic and synthetic distinction: it might be analytically true that humans would follow something like

Rawls' two principle of justice if they were rational, but to show this goes beyond the scope of Rawls' book. I hope, however, to have raised the issue of the grounding of reason with sufficient urgency to show that metaphysical neutrality is not an option for a theory of ethics.

CHAPTER III

HEGEL'S AND WILLIAMS' CRITIQUES OF THE KANTIAN POSITION

In this chapter I look at what it would take to resolve the problem of normativity which I have so far developed with regard to Kant and Rawls. To do this, I would like to turn to two influential criticisms of the Kantian position, those of Hegel and Bernard Williams. These criticisms are not exactly the same, but they do share a fundamental skepticism that the moral can be divorced from the non-moral realm. If Hegel and Williams are right and the non-moral and the moral cannot be separated, we must inquire into what it would mean to see them as one instead of dividing them up the way Kant and Rawls have. Seeing the moral and the non-moral as one would mean showing how normativity pervades all realms of thinking and acting. It would be, to take up Prauss' criticism from chapter one, to show that there is a way in which thinking is already acting.

In addressing this point, this chapter will function as a transition point between the two previous chapters which sought to examine what a theory of practical reason might look like with and without a metaphysical basis. The next chapters will deal with the need to refigure what is meant by a 'metaphysical' justification for moral conduct. In this sense, the chapters coming after this will have to contend with the question of what it means to work through and alter the question of universality itself with regard to moral conduct.

I will present Hegel's criticism first, both for historical reasons and because his criticism of the reciprocity thesis shows that no stable and strong sense of universality can be made available in the Kantian argument. Williams' criticism, by contrast, focuses more on the question of whether there can be any sort of universal justification for action at all.

The Critique of the Categorical Imperative I: The Formality Charge

Hegel's critique of Kant's ethics is usually characterized as the charge that the categorical imperative is empty. At a deeper level, however, this charge amounts to skepticism about the authority that action generated by the categorical imperative would have. I will thus argue that Hegel's criticism of Kant really claims that the Kantian theory suffers from a disjunction between the meta-concept of the categorical imperative and its application. If such a disjunction exists, then whatever content the categorical imperative does provide is irrelevant to the actual conditions to which it is meant to be applied and hence without authority to instantiate the moral action. This means, as Hegel puts it, that any action the agent does decide on will be necessarily subjective—bound only by the authority of the agent's desires and therefore not autonomous.

Hegel's critique of the categorical imperative falls into two parts. The first part concerns the contention that the categorical imperative is empty because it cannot generate principles for action. This is the first order issue of whether the categorical imperative can generate contentful principles. The second part of the

critique concerns the authority which the contentful principles generated by the categorical imperative (if there are any) have for the justification of the principle of practical reason in general. This second-order concern must deal with the question of how any content at all can be found in the noumenal realm given Kant's total exclusion of nature there. In other words, do reasons for action stem purely from the noumenal realm or do they come from some dialectic of reason and nature, as Hegel seems to propose?

I will deal only briefly with the first-order issue since it is well known. Hegel makes two main criticisms of Kant's conception of the maxim. The fundamental objection to Kant is stated programmatically: GPR §132:

The *right of the subjective will* is that whatever it is to recognize as valid should be *perceived* by it *as good*, and that it should be held responsible for an action— as its aim translated into external objectivity— as right or wrong, good or evil, legal or illegal, according to its cognizance of the value which that action has in this objectivity”.⁹⁰

The main point here, as before, is that the subjectivity which defines the moral standpoint renders it arbitrary against the objective standpoint of reality. As Hegel explains a little later, the problem is that the ‘ought’ which the moral point of view entails cannot generate usable principles because the ‘ought’ is a general injunction to do good and one cannot do good in general. But Hegel also praises

⁹⁰ Hegel. *Foundations of the Philosophy of Right*. §132 Henceforth cited as GPR followed by paragraph number. “Das Recht, nichts anzuerkennen, was Ich nicht als vernünftig einsehe, ist das höchste Recht des Subjekts, aber durch seine subjektive Bestimmung zugleich formell, und das Recht des Vernünftigen als des Objektiven an das Subjekt bleibt dagegen fest stehen.”

Kant's recognition that one is free only if one does one's duty. Hegel sees Kant's great insight as having shown that duty and freedom are really one in the sense discussed earlier— that freedom requires us to adhere to certain principles. The dispute here is thus about the ability of the categorical imperative to generate contentful principles.

Hegel's first order criticism of the categorical imperative is that it cannot yield concrete ethical principles because it merely takes the form of the law of non-contradiction. Thus Hegel contends that it is just as consistent to will private property as to will no private property.⁹¹ This objection targets the consistency in conception version of the categorical imperative-test. The objection, however, can be dispelled on the grounds that, in the deposit example from the *Groundwork*, what is in question is not so much the willing of private property as the notion of trust between people, upon which the practice of the deposit is based.⁹² It might thus be wished that trust continue to exist while I abuse it, but this cannot be willed in a strict sense of willing: I can wish that I might have my cake and eat it too, but I cannot will both at the same time, for this cannot be logically conceived of. Though Hegel's concrete criticism of this example can be turned back by clarifying the maxim, the point that the maxim is itself necessarily concrete and as such subject to misunderstanding stands.

⁹¹ The short-story "The Black Sheep" by Italo Calvino makes this point very well. In the story, the peaceful life of a town of thieves, who nightly steal from each other, is disrupted by the arrival of an honest man who refuses to steal, though he is stolen from. This, in a Rousseauian manner, leads quickly to some attaining more material possessions and others less. The point seems to be, as in Hegel, that it was the consistency that mattered, not the principle itself.

The second objection, articulated most clearly in the natural law essay, concerns the idea that the categorical imperative aims at its own extinction.⁹³ This objection claims that to will the end of poverty and to make this one's principle of action aims at the eradication of the need for this principle and thus of morality more generally. According to Korsgaard and Wood, however, the problem lies in the correct formulation of the maxim. The maxim should not be 'end poverty' since this is a state that might, conceivably, be eradicated, but rather 'help those in need', a more general formulation, or even, 'do good to others', which is surely not a maxim that can ever become redundant.⁹⁴

The problem with Kant's categorical imperative is thus not that it does not permit us to determine which maxims are permissible and which are not, but that Kant does not tell us how to formulate maxims, since a given situation might be formulated in a variety of ways. As Wood points out, claiming that we should consider only those morally salient issues in stating our maxim only restates the problem since it is precisely moral salience which is at stake in the maxim.⁹⁵

I want to interpret these criticisms in terms of the problem of synthesizing inclinations and principles. Hegel's critique claims that because the application of the principle to the inclination is in a sense arbitrary, any duty to act which results from this synthesis will itself be arbitrary. Hegel thus argues that without a better understanding of how reason and inclination relate no such thing as rational

⁹² See Allen Wood. *Hegel's Ethical Thought*. 157-58.

⁹³ Hegel. "Über die wissenschaftlichen Behandlungsarten des Naturrechts". (1802).

⁹⁴ Allen Wood. *Hegel's Ethical Thought*. 160.

⁹⁵ Wood. *Hegel's Ethical Thought*. 161.

action is possible. What we need, in other words, is a better understanding of how the maxim itself comes into being and how maxims form as a bridge between the self and nature.

The Critique of the Categorical Imperative II: The Failure of the Reciprocity Thesis

Though it seems possible to generate some kind of principles out of the categorical imperative, this does not resolve the question of whether these principles are sufficient to actually allow us to accomplish particular acts. The question Hegel raises about the underdetermination of the moral principles remains threatening to the Kantian project until we can determine whether the categorical imperative has the power to justify and motivate action according to the principles derived through the categorical imperative-test.

This question leads us all the way back to the subject matter of the first chapter, in which I examined Kant's attempt at proving the reality of pure practical reason. The conclusion reached there, even by Kant, was that the attempt at a proof that reason and morality imply each other was a failure. I would now like to examine whether there might yet be a way out of this problem given a different reading of Kant. To this effect, we should examine Allison's and Korsgaard's defense of the fact of reason doctrine.

In chapter 1, I dealt chiefly with the problem Kant had in showing that we humans are rational creatures, i.e. the second part of the proof of morality. Here we must deal with the first part, the claim that, as rational creatures, we are subject to the moral law. This question is central to Hegel's more fundamental

critique of Kantian morality, especially to the emptiness charge. For if it cannot be shown that reason can justify itself, that is, provide principles of action which are purely rational (the categorical imperative), then it cannot be shown that autonomous reflection can give us reasons to act.

Allison has called Kant's attempt to show that reason can both justify and motivate rational creatures the reciprocity thesis. It claims that morality and autonomy entail each other. The general statement of the reciprocity thesis goes as follows: (1) As a 'kind of causality' the will must be law-governed or determinable according to some law (a lawless will is an absurdity); (2) as free, it cannot be governed by the laws of nature; (3) it must therefore be governed by laws of a different sort, namely, self imposed ones; and (4) the moral law is the required self-imposed law.⁹⁶ The point is that for Kant freedom is not only a necessary condition for morality, but also a sufficient condition.

This line of reasoning has been attacked by many but I will examine it in the formulation of Allen Wood who shows that the argument rests on an ambiguity in the notion of universality employed by the categorical imperative. Wood claims that, though Kant can rightly claim that the principle of practical reason must be universal, this universality entails only a universality in applicability but not in both applicability and concern. That is, I might, as the rational egoist does, think that it would be rational for everyone to act as I do, but also wish that not everyone actually does. This means that autonomy might only

⁹⁶ Allison. *Kant's Theory of Freedom*. 202-204.

yield the hypothetical imperative but not the categorical imperative.⁹⁷ And this means that autonomy might be necessary but not sufficient to establish morality.

Let us examine the problem in a little more detail. The reciprocity thesis is meant to stop a regress of reason-giving. Charles Taylor puts the problem well: “If freedom is to renounce all heteronomy, any determination of the will by particular desires, traditional principles or external authority, then freedom seems incompatible with any rational action whatsoever. For there do not seem to be any grounds of action left, which are not wholly vacuous, that is which would actually rule some actions in and others out, and which are not also heteronomous.”⁹⁸ Having renounced all forms of nature like inclination (internal) or contingency (external), it looks like there might not be anything to put an end to the regress, that is, nothing could have enough content or authority to justify the maxim we should act on and thus motivate action.

Allen Patten’s treatment of the issue in Allison and Korsgaard is excellent and I follow his analysis here.⁹⁹

Kant’s argument starts with the weaker sense of universality and ends with the stronger sense, which means that there is a gap to be bridged. Allison has argued that the gap can be bridged by appealing to transcendental freedom, the idea that freedom consists in complete ‘independence from everything empirical and hence from nature generally’. Transcendental freedom refutes the threat of rational egoism by showing that the latter doctrine relies on a material determining ground of the will— rational egoism presumes an end. Allison’s defense thus amounts to saying that

⁹⁷ Allen Wood. *Hegel’s Ethical Theory*. p. 135. See also David Gauthier. “The Unity of Reason; A Subversive Reinterpretation of Kant”.

⁹⁸ Charles Taylor. *Hegel*. 373. Cited in Allen Patten. *Hegel’s Idea of Freedom*.

⁹⁹ Allen Patten. *Hegel’s Idea of Freedom*. 91-93.

principles that purport to be universal in the first sense but not in the second, are not universal at all. But though Allison may be right to say that the principle of rational egoism could only be deemed reasonable in the first place because of certain presupposed ends, it still might be the case that the same is true of the categorical imperative. Why is the categorical imperative deemed to be intrinsically rational and rational egoism not?

In “Morality as Freedom”, Korsgaard makes explicit two assumptions that might well underlie Allison’s argument, namely that (1) a spontaneously or transcendently free agent— an agent who is independent of all authority— has as a highest order reason or ‘incentive’ the preservation and promotion of his spontaneity, his independence of all empirical conditions, and (2) that someone who has this incentive would reject, upon deliberation, every principle of action except the principle of acting on only those maxims that could at the same time be willed as universal law.

Kant’s justification for the second claim is: “The moral law does not impose a constraint on the will; it merely says what it has to do in order to be an autonomous will at all. It has to choose a law.”¹⁰⁰ According to Patten, the ambiguity in ‘choose a law’ here mirrors exactly the ambiguity of ‘universality’ in Allison’s argument. Thus it might be true that the will must choose a law for itself, but this does not mean that the will must choose a law which others must follow as well. Patten thus concludes that Kant’s argument for the reciprocity thesis fails because, though valid, it relies on a faulty premise.

¹⁰⁰ Christine Korsgaard. “Morality as Freedom”. 166.

The Problem of Intersubjectivity in Kant and Rawls

The inability of Kant's theory to prove the objectivity of reason means that for Kant there is no way of establishing morality. The threat remains that what is rational for me might be rational *only* for me. This leaves Kant without a theory of intersubjectivity, for his theory of sociality depends on the objectivity of reason, which has not been established.

Rawls too attempted to provide a theory of intersubjectivity based on the rationality of the individual subject. In *Kantian Constructivism*, Rawls argued that the rational (the hypothetical imperative) must be framed by the reasonable. This is so because, from the perspective of the agents for whom the results of the deliberation in the original position are supposed to be normative, it makes sense only to take these results as embedded in certain commitments like respect for persons which we (as really existing agents) already have.¹⁰¹ Thus the instrumental rationality of the hypothetical imperative is overlaid with a normative perspective of respect for persons. Rawls argues, in effect, that the reasonable and the rational are one and the same principle. Rawls thus also seeks to move from the weak autonomy thesis to the strong one.

The problem with Rawls' argument, however, is that the argument must assume the principle of pure reason which the argument is itself meant to establish. This leaves Rawls' theory open to the charge that it, like Kant's argument, never moves beyond the hypothetical imperative in the sense that my reflection might commit *me* to certain acts but not others as well. We see this

problem resurface in the reflective equilibrium: the reflective equilibrium, I concluded earlier, is only ever able to tell us how much agreement there already happens to be with those around me. But the conclusions of others are not normative for me, because I am only ever committed to my own conclusions. There is no objectivity of pure reason which can guarantee that my conclusions are the same as yours.

The problem with Kant and Rawls' arguments is that they rely on the principle of pure practical reason to establish the strong autonomy thesis (which states that freedom or morality exists) and then turn around and infer from the principle of reason that morality or intersubjectivity exists. The argument, in other words, is circular because there is no independent proof of the existence of pure practical reason.

This is a powerful critique of the Kantian position and we will spend the rest of this dissertation examining its ramification. Before we begin looking for a way to meet this objection, however, I would like to bring in one more formulation of the problem. I propose now to turn to Bernard Williams' critique of the Kantian position.¹⁰² Williams' formulation has two advantages over Hegel's formulation. The first is that the answer to the Kantian problem of the justification of morality which we will be pursuing in Korsgaard's work is actually in part addressed to

¹⁰¹ See also Rawls. "Themes in Kant's Moral Philosophy". 503-504.

Williams' concerns and so does not force us to translate quite so much from the language of recent Anglo-American philosophy to that of 19th Century German philosophy. The second, perhaps more important, point is that Williams bases his objections to Kant on a theory of person and society which can more readily be summarized than Hegel's theory can. By examining Williams' view of the historical situatedness of the person we will thus not have to beg the question quite as often as we would in considering a scaled down version of Hegel.

A Reorientation of the Normativity Question: Justification and Motivation

As a way of returning to Williams' account, it would be good to see where Hegel's critique has taken us. As I said at the outset, this dissertation is not principally concerned with the letter of Hegel's critique of Kant but rather with the ways in which certain questions arising from this critique have persisted in more recent scholarly debates between Kantians and Hegelians.

The question I take as fundamental is what the failure of the reciprocity thesis means for future attempts to ground normativity. The criticism of the Kantian position according to which my reflections might motivate *me* to act but not others destabilizes what we think of as a philosophical or moral justification for action. This suggests a dilemma: we must avoid concluding both that moral

¹⁰² I refer to the Kantian position since Williams' critique is aimed just as much, perhaps more, at contemporary writers like Rawls than it is at Kant himself. Indeed, Williams refers to the texts of post-Kantian writers far more often than to Kant himself.

reflection is merely justifying but not motivating (because it is too abstract) and that it is motivating but not justifying (because it is not universal).

For our purposes, the alternative most pressing is the latter because I am assuming throughout that action is indeed possible. We are here concerned with the question of which actions are moral in the sense that they are universally permissible. The Hegelian criticism must yield a new way of thinking through the universalizability requirement which preserves both the need for principles to motivate us and the idea that they should be in some sense imaginable as universally permissible. If there is no transcendental realm in which the laws of morality are in some sense given to us through reflection, we must reconsider what universal justification can give us. Reflection might only ever be able to give us in some sense temporary and revisable principles according to which we can act 'morally'.

With this in place, we should turn to Williams' critique of the Kantian position. Williams seeks to bring to bear the finite nature of our process of reflection on the question of universal justification for action.

Williams' Critique of the Kantian Position

Williams makes two general criticisms of the Kantian position. The first point concerns Williams' claim that reflection and detachment are not the same thing. That is, the ability to detach oneself from one's commitments is not the

same as being able to reflect on them.¹⁰³ The second point is an elaboration of this to the effect that, since one cannot detach oneself from one's commitments, it is wrong for a system of impersonal morality to demand that one do so.

Let us start with the first point, which I shall call the internalism thesis. It has two sub-points. The first concerns the idea that reasons for doing something must always be internal to the agent in order to motivate him. The second point is that the first claim, at least for Williams, includes to some extent the Humean assumption that desires motivate directly. Let me elaborate.

In his famous paper "Internal and External Reasons" (in Scanlon's useful summary):

Williams distinguishes between two ways of interpreting a statement that a person "has a reason to ϕ" (where ϕ stands in for some verb of action). According to the first interpretation, the statement implies that the agent has some motive— that there is something that matters to him or her— that will be served or furthered by ϕ-ing. If the person in fact has no such motive, then on this interpretation the claim is false. The second interpretation includes no such condition, so on this interpretation it can be true that A has a reason to ϕ even though ϕ-ing would not serve or further any aim or value that matters to the agent. Williams refers to reason claims under the first interpretation as "internal reason statements" and refers to those under the second as "external."¹⁰⁴

This shows that Williams is not a skeptic about reasons, but that he understands reasons differently than writers in the Kantian tradition do. Here I

¹⁰³ See also Allison. *Kant's Theory of Freedom*. 194-95. Allison argues that Williams' central criticism of Kant amounts to the charge that Kant has conflated the theoretical and the practical standpoint in arguing that reflection and detachment are one and the same.

¹⁰⁴ Thomas Scanlon. Appendix. *What We Owe Each Other*. 363.

follow Korsgaard's distinction between content and motivational skepticism.¹⁰⁵

Korsgaard argues that Williams' skepticism about the ability of pure practical reason to motivate is due not so much to the idea that pure practical reason cannot, in principle, motivate, but rather to Williams' belief that it has not been shown to do so. This is due to Williams' sub-Humean conception of reason.¹⁰⁶

The sub-Humean model supposes that ϕ -ing has to be related to some element in the motivational subset S as causal means to an end. But here reasons are construed rather more broadly than on the Kantian view, for according to Williams one can have many reasons, as well as reasons not to follow these reasons. Reasons may be imaginatively interrogated and in doing so one might find that what one thought was a reason really is not one, or one might find that one has reasons to do things one did not think one had.

There is thus no question that in Williams' sub-Humean model of motivation, there is deliberation of some sort. Indeed, deliberation seems to run rampant. The set from which one draws reasons to deliberate about includes all psychic states which are part of what makes a character, including pleasure, "dispositions of evaluation, patterns of emotional reaction, personal loyalties: and various projects, as they may be abstractly called, embodying commitments of the Agent."¹⁰⁷

¹⁰⁵ Christine Korsgaard. "Skepticism about Practical Reason".

¹⁰⁶ The thesis is 'sub-Humean' because, as Williams argues, it does not contain all the details of Hume's own account.

¹⁰⁷ Williams. "Internal and External Reasons". 105.

Williams has thus extended the set of motivations for action from mere direct desires to a whole host of other interests and commitments. In doing so Williams has also diluted the force of his internalism to the extent that customs can now be included, giving the agent's action an intersubjective grounding. The point to bear in mind, however, is that for Williams what really counts here is that whatever the reasons for my actions are, they are *mine* and therefore not universal or objective. This means simply that you cannot give me reasons for acting. They might become reasons for acting only if I adopt them by making them my own in deliberation.

We have already examined this point in a slightly altered form in the failure of the reciprocity thesis in Kant. It was the idea that though it might be true that your projects are normative for you in just the same way my projects are normative for me, this does not require me to take the fact that you have chosen your projects the way I choose my projects as a reason not to interfere with your projects.¹⁰⁸ The argument thus repeats the charge that normativity might not be sharable.

It is important to see that while the internalism thesis appears to undermine the Kantian project of arguing that a universal morality is motivating, the thesis reinforces the point that something like the weak autonomy thesis is true, that we are motivated by our own deliberation. This claim, of course, still

¹⁰⁸ Williams. *Ethics and the Limits of Philosophy*. 59-62. In formulating this criticism, Williams is also referring to Alan Gibbard's argument to the same effect. See Alan Gewirth. *Reason and Morality*.

leaves open what, exactly, such non-moral or morally neutral deliberation means for Williams, and also what it means for post-Kantian writers like Korsgaard. To put the question in terms of synthesis, it now appears that some type of deliberation is able to synthesize reason and nature such that action can result from it. The question now becomes, what is the nature of this deliberation? Though Williams is vague on this point, Hegel and Korsgaard both have considered views about what constitutes both morally neutral and moral deliberation which we will turn to in this chapter and the next.

Before we turn to the process of deliberation, we should, however, take up Williams' second criticism of Kant, the idea that the impersonal moral standpoint violates the integrity of our character. This criticism is really an elaboration of the argument that we are motivated by our own particular subset of available interests and desires and that to force us to adopt reasons that are not ours is tantamount to coercion.

I would like to briefly look at two different arguments Williams gives to justify this point. In "Persons, Character and Morality" Williams argues against Rawls' idea that we owe equal allegiance to our current selves and our future selves and must thus seek to maximize the benefit to our present and future selves equally. Against this Williams argues that the correct perspective from which to consider one's plans and goals, one's ground project as Williams puts it,

is *from now* since it is only the current perspective that motivates me to go on at all, that is, that keeps me from suicide.¹⁰⁹

In “Moral Luck”, Williams draws the argument out a little more by suggesting that the deliberation process we engage in to make certain life altering decisions is not universal because it depends for its justification on a future outcome that is necessarily not yet known. He argues that the justification of Gauguin’s decision to leave his family and go to Tahiti can only be ascertained at the end of the project. The futurity of the justification (to Gauguin— there may never be a justification which his family would accept) introduces the factor of moral luck in the sense that it depends on a certain set of contingencies that determines whether the project succeeds.

Williams introduces the notion of agent-regret, a type of regret felt only by the agent involved, which centers on the desire for things to have turned out differently, but does not necessarily include the desire to have deliberated differently. Williams uses the tension between the success or failure of a life project with regard to the luck and/or correct deliberation involved in it to argue that what ultimately matters to the agent is not the correct deliberation but the outcome. For Williams this means that there is a disjunction between the rational deliberation in pursuing our ends advocated by the impartial (moral) perspective, and the luck or unluck that the success of our particular ground plan requires in order for it to be genuinely ours and to count as success for us. In this sense, it was Gauguin’s bad luck to be the type of person who had to make the decision to

¹⁰⁹ Williams. “Persons, Character and Morality”. 12.

break with morality in order to pursue his ground project. And since, by the internalism thesis, Gauguin's character is motivated by his particular set of interests and concerns, there is something wrong with a moral scheme that seeks to divorce us from our fundamental projects. Williams concludes that morality must thus be seen as one value among others or be refigured in order to include individuality more thoroughly.

Williams' critique of morality thus seems to suggest that the realm of normativity should not be seen exclusively in terms of either moral duty or (morally neutral) permissibility. Normativity goes all the way down, as the internalism argument shows. This means that for Williams we should start our ethical inquiry with something more akin to Socrates' question: 'how should one live'.¹¹⁰ This question does not presuppose a stark distinction between the moral and the non-moral but rather a continuum of sometimes conflicting normative demands, all of which we must seek to meet. Failure to meet these demands will leave us with the sort of regret which Williams argues plays a central role in accomplishing or failing to accomplish our projects.¹¹¹

¹¹⁰ Williams. *Ethics and the Limits of Philosophy*. 1. Williams, however, expresses some doubt that the impersonal way of formulating even this question, might lead us to a false understanding of ethics. *Ethics and the Limits of Philosophy*. 4.

¹¹¹ See also Martha Nussbaum. *The Fragility of Goodness*. pp. 1-84. Nussbaum argues that the emphasis on the impersonal Christian and Kantian standpoints has actually obscured our understanding of Greek thought, since Greek thought was in part developed in tragedy and not only in abstract ethical theory. Nussbaum thus argues that the tragedians (she uses *Antigone* and *Agamemnon* as examples) show that in genuine tragedy—having to choose between the survival of your army and the survival of your daughter, for instance—there is no decision that will leave

In emphasizing the Socratic question over that of duty, Williams is in accord with Hegel's criticism of Kant as well. Neither Hegel nor Williams deny that there are such things as more narrowly moral commitments, but both emphasize that because the moral can sometimes lead to an impasse, we must consider ways in which to re-imagine the moral, making it more inclusive.¹¹² In this sense, both Hegel and Williams emphasize the ethical or *Sittlichkeit*, over the moral.¹¹³ But depending on how one reads Hegel, he is either much more optimistic about the possibility of reconciling our desires with those of others in our society or only a little more so than Williams who does not hold out much hope for an eventual resolution to the ethical problem. We should also note that the position of Williams and Hegel bring us back to an objection to Kant raised in the first chapter by Prauss, who suggested that Kant had really made a mistake

you clear of blame. The failure to acknowledge this, and not the decision to sacrifice his daughter Iphigenia, is what makes Agamemnon's actions blameworthy. It should be noted, however, that Kant does not emphasize blame but rather responsibility. Blame, it appears, only makes sense from within a thick conception of society such as is perhaps no longer apparent in Kant.

¹¹² By calling this a criticism of Kant, I do not mean to suggest that Kant does not himself offer ways of getting beyond such impasses. Kant's philosophy of history is centrally concerned with just this issue. However, the apparent division between Kant's ethics and his philosophy of history invites the belief that there is something about 'morality' construed in the individual sense, that Kant wants to hold up against other more social theories.

¹¹³ Williams notes that the moral is a subset of the ethical which emphasizes certain aspects over other and says that we should treat this idea with particular skepticism. *Ethics and the Limits of Philosophy*. 6. Hegel's discussion of morality (and in particular the Kantian view) in the *Foundations of the Philosophy of Right* leads to the dilemma of empty personal conviction which is at odds with the rest of society.

by trying to prove autonomy by linking it to morality directly rather than exploring other possibilities of understanding autonomy as normativity.

A further point to be made here about Williams' and Hegel's critique of Kant and the Kantian position is that the critique concerns the special status Kant gives to normativity as moral rather than the status of normativity in general. Both Williams and Hegel believe that normativity is rather more general than the traditional Kantian position as well as theories of rational choice or egoism make it out to be. The question here is whether there is a way of conceiving of normativity in social terms without Kant's emphasis on the impersonal standpoint. To put it in the language used earlier: is there something about deliberation which is historically specific and personal, and thus capable of motivating us and yet aspires to universality without claiming to be universally legislating?

Lastly, there is the question of moral luck itself, that is, the question of how (or whether) the individual can be reconciled to the prevailing ethical schema. As the Gauguin example makes clear, there is a certain recalcitrance of individual commitments which Williams diagnoses. This is a question which Korsgaard does not really deal with, but one which, I hope, Hegel can shine some light on. This question too, is the question of the social which we have been tracking so far. In the absence of Kantian rigorism, it now appears that we might lose universality altogether, and that being moral may not be an achievement at all, but a mere contingency, a happy coincidence of the individual commitments of different agents. But taken together with Williams' argument about the irreducibly

subjective character of normativity, this is a picture that Kant, Korsgaard and Hegel want to reject vehemently.

Reformulating the Questions after Hegel and Williams

We have now arrived at a critical point in the development of the argument since we have come to a better understanding of the question to be answered by the two remaining writers to be considered here: Korsgaard and Hegel. There are two main problems that need to be addressed in order to provide at least the beginning of a satisfactory resolution to the problems raised so far. The first issue is that of internalism and reflection. Here we will need to see whether a conception of reflection can be developed which is able to successfully synthesize the thoughts and intentions of the agent with the contingencies of the world. This reflection will thus have to be normative, a reason for action, for the agent who conceives of it. This will be the subject of chapter four. The second question we will need to answer is whether, if we can develop a theory of normativity in the above way, this normativity will in some way bind not only the agent to her actions, but agents to each other as well. This will be the subject of chapter five. We will have to see whether the new conception of normativity will have intersubjective validity.

In taking up the question of intersubjectivity, this chapter will also try to make some inroads into the status of moral luck. For if reason is indeed intersubjectivity, the issue of moral luck takes on a rather less damaging character since it then might be (merely!) the lack of appropriate historical

development that condemns the subject to living at odds with the prevailing social norms, rather than a fundamental incompatibility of commitments. The question of moral luck forces us to examine the relative accessibility of intersubjectively harmonized standards of action in the modern world. The response to moral luck, as Rawls keenly appreciated, is always more justice.

If it is indeed possible to reformulate the Kantian question with regard to the criticisms of Williams and Hegel, using two writers who are not often found together in the same work let alone in the same argument, I will take this to have shown that there is much greater similarity between 19th Century post-Kantian philosophy and late 20th/ early 21st Century post-Kantian philosophy. The proof that this is so will, of course, only be evident at the end of the argument itself. However, I am, I believe, at least justified in suspecting that Hegel and Korsgaard provide a way out of the problems posed for Kant by Hegel and Williams. I thus disagree with the frequent lumping together of Hegel with Williams though as the earlier part of the chapter sought to show, they level the same sorts of criticisms at Kant.¹¹⁴ The difference is that Hegel sees a way out of Kant which is sympathetic to Kant while Williams does not.

Prospects for a Reformulation of the Kantian Project

Let me now say a little about how I propose to give an answer to the above question using Hegel and Korsgaard. Building on Korsgaard's strong

¹¹⁴ See Allison. *Kant's Theory of Freedom*. ch. 10.

Kantian credentials, I will examine her reinterpretation of the Kantian process of reflection, which avoids at least the first Hegelian objection, that Kant's maxims cannot motivate any particular action in the world. Korsgaard's argument will also conform to Williams' internalism requirement by being absolutely subject dependent.

But this still leaves us only with a notion of normativity which is subject specific and a second argument will be required in order to show that normativity is more than egoistic. To this effect, I will examine Korsgaard's idea that reason is public and that we therefore, at least in principle, can have shared aims and projects. While Korsgaard provides only brief arguments for the second point, her analysis of action seems to bridge the gap between individual normativity and more general normativity. Her theory of action centers on the idea that, in acting, we set about doing things according to constitutive standards of action which have both technical standards and social standards, corresponding to the hypothetical imperative and the categorical imperative. The fact that when building a house, we seek to conform both to socially and technically shared standards, I will argue, gets us half way to public reason. The other half of the way is provided by an argument which seeks to show that we cannot be just to ourselves and not also at least feel pressure to be just to others.

In bringing Hegel into the discussion, I will argue that Hegel and Korsgaard share the idea that reflection on desires must be conceived of as rational and furthermore that both writers also share the idea that reason must be conceived of as public or intersubjective. I will further argue that Hegel's theory of

institutions provides an important extension of Korsgaard's argument about the constitutive standards of action, since, in Hegel's view, institutions unify the subjective projects of individuals with the more general and public projects of society itself. I will thus try to expand on Korsgaard's theory of action by supplementing it with Hegel's view of the subjective and objective elements of the state.

Hegel's view, I will argue, expands Korsgaard's notion of reflection to include historical and social practices which aspire to be universal but do not pretend to be so already. In this sense, social practices are the measure of our current responsibilities to each other but at the same time are subject to revision as new standards which are more inclusive and more just emerge through the process of reflection. This thought, I will argue, lies at the heart of both the Kantian and the Hegelian enterprise. But the Kantian ethical enterprise goes wrong when it insists that metaphysical objectivity is at every turn accessible to the individual agent, no matter her historical position. This claim is, of course, itself in tension with Kant's own philosophy of history where Kant develops a dialectic of historical progress in which each individual must seek to overcome her place in history but where pragmatic historical conditions are also taken into account.

CHAPTER IV

KORSGAARD ON SELF-CONSTITUTION

In this chapter I would like to present Christine Korsgaard's theories of reflection and action. Together, these two theories will show that humans are normatively motivated and that our actions cannot occur except under a unified process of reflection which, Korsgaard argues, actually constitutes the self. This will then lead us to Korsgaard's theory of action which is meant to show that reflection constitutes paths of action which are normative for us. This normativity is given an objective content by Korsgaard's argument that actions have their own standards. The discussion of reflection is meant to be neutral about the subject of reflection in the sense that this chapter does not seek to answer the question of whether the reasons that derive from reflection are private or public, subjective or intersubjective. That question will be dealt with in the next chapter. Nevertheless, when discussing examples of standards of action which arise in the discussion of standards of action, I will try to remain neutral about the ethical or moral content of these standards. This neutrality is part of Korsgaard's larger project of avoiding a conception of normativity which is explicitly linked to a morality.

In examining Korsgaard's argument about reflection and action we will have to keep in mind the two criticisms Hegel and Williams make of Kant.

Korsgaard will have to show that reflection is not empty, i.e. that it leads to action, and that action stems from the desires of the subject itself in order to at least partially deflect these criticisms. The former will satisfy Hegel's criticism that the categorical imperative is empty and the latter will satisfy Williams' internalism requirement.

The underlying themes of this chapter are threefold. The bulk of the chapter will be devoted to showing how Korsgaard's theory of the self is developed out of the theory of reflection. In this, I take her to be quite close to the basic tenets of German idealism according to which the self's consciousness of itself is the origin of both knowledge and action. By restating the account the way I do, I am also trying to show that Korsgaard's theory of reflection is not all that different from Hegel's account of the origins of intersubjectivity as we will see it developed in chapters six and seven. Lastly, I want to show in this chapter that egoism is not a threat to this self-positing conception of the self because egoism's central thesis, that we are really only desire-driven creatures, is insufficient to even develop a sense of self.

The Question of Moral Character

The question of moral character is, for Kant and for Korsgaard, centrally one of the possibility of reflecting on one's situation through the employment of formal reason. This means that morality is a question of becoming independent of material or natural concerns. This idea, however, has often been challenged. In *Ethics and the Limits of Philosophy*, for example, Williams raises the question

of why, given one's substantive personal commitments, it is even useful to imagine oneself as separate from them. Williams writes: "[In ethics] we are concerned with what any given person, however powerful or effective he may be, should reasonably do as a rational agent, and this is not the same thing as what he would reasonably do if he were a rational agent *and no more*."¹¹⁵ That is, Williams doubts that it is even possible to intelligibly represent oneself as devoid of commitments in this way and that if we were able to represent ourselves as stripped of our commitments, what we would do might be completely different from what we would do committed as we are. Yet this is what the Kantian maxim seems to demand. Behind this lies Williams more fundamental worry that without the commitments we do have, we would not have any commitments at all. Distancing ourselves from the commitments we have does not change this.

Korsgaard argues that the opposite is true, namely that philosophical reflection does not leave everything just where it was.¹¹⁶ This, as I have said, is the core of the disagreement between Williams and Korsgaard. Let us now turn to a first formulation of why Korsgaard thinks this is so. Korsgaard begins from the fact that we are self-conscious beings and that the outcome of our reflection on the world gives us our identity, that is, our commitments and obligations. Our commitments and obligations give us our character. But since our commitments and obligations are always changing, our identity does not rest on any single

¹¹⁵ Williams. *Ethics and the Limits of Philosophy*. 63.

¹¹⁶ Korsgaard. *The Sources of Normativity*. 119.

substantive set of commitments but merely on the fact that we have *some* commitments.

So, for Korsgaard, the fact that you are governed by *some* practical identity is not contingent. It is your humanity itself. Korsgaard writes: It is because we are human that we must act in the light of practical conceptions of our identity, and this means that their importance is partly derived from the importance of being human. We must conform to them not merely for the reasons that caused us to adopt them in the first place, but because being human requires it.¹¹⁷

Our humanity provides the normative glue which makes the contingencies of the world into a practical identity which we feel we must maintain even in the face of adversity. Humanity, taken as the ability to reflect on one's circumstances, is thus the reason on which we act. From this Korsgaard concludes: "Since you cannot act without reasons and your humanity is the source of your reasons, you must value your humanity if you are to act at all."¹¹⁸ This is a transcendental argument: Rational action exists, so we know it is possible. How is it possible? And then, by the course of reflections in which we have just engaged, I show you that rational action is possible only if human beings find their own humanity to be valuable.

¹¹⁷ Korsgaard. *The Sources of Normativity*. 121.

¹¹⁸ Korsgaard. *The Sources of Normativity*. 123.

But rational action is possible, and we are the human beings in question.

Therefore we find ourselves to be valuable. Therefore we are valuable.¹¹⁹

Korsgaard's argument is directed against Williams' claim above that detachment does not mean reflection. Recall that for Williams, though it is quite possible for us to distance our selves from our circumstances in thought, such a distance does not furnish any sort of guidance for action because there is no content upon which to base a judgment. But by arguing that our conception of humanity underlies our conception of what it means to have a practical identity, to have commitments in general, Korsgaard suggests that it is really only *because* we have an intrinsic value that such a thing as practical identity or Williams' ground project is possible at all. This means that reflection is constitutive of character or practical identity. This means that we have established that my projects are normative for me. We could also put it thus: normativity arises from the structure of our self-consciousness rather than from our intra-worldly commitments.

And not only that, the fact that our projects have value to us also means that they have value to others since humanity, the ability to choose our projects rationally, is something each person has. In this last step which moves us to intersubjectivity proper, Korsgaard takes up Williams' skepticism about the idea that what you take to have value for you (your ground project) must also be valuable for me. Why must your ground project be valuable for me? Korsgaard

¹¹⁹ Korsgaard. *The Sources of Normativity*. 124.

argues that since all value has the same origin in our humanity, we must respect that source of value and everything that springs from it.

This last step has come in for criticism. The gist of one criticism, voiced by G. A. Cohen, is that practical identity is not sufficiently well connected to universality or reason to be normative.¹²⁰ Cohen argues that while Kant tries to show that morality comes from reason, Korsgaard only bases her conception of morality on practical identity. Practical identity, however, is contingent in important ways which means it is not a stable enough concept to underwrite humanity itself. Practical identity, Cohen argues, may not have anything at all to do with the moral law, but might just be a collection of impulses. Cohen's objection thus amounts to the worry that without Kant's metaphysical notion of reason, practical identity might be merely heteronomous action all the way down, that is, there is no obvious connection between believing something to be valuable and it being valuable without the guarantee of the unity which only reason can provide.

It may be that Korsgaard is responding to such criticism by dropping the notion of practical identity in the *Locke Lectures*. The important thing to take away from the discussion of value in *Sources* is that Korsgaard takes seriously Williams' challenge to give an internal explanation for reasons. We have now

¹²⁰ G. A. Cohen. *The Sources of Normativity*. Lecture 6. "Reason, humanity, and the moral law." For another version of the concern, see Eric Watkins and William Fitzpatrick. "O'Neill and Korsgaard on the Construction of Normativity". They argue that the problem with Korsgaard's approach is that value is not basic but only willed. This seems to make it contingent, in some sense. 361

seen Korsgaard try to give a truly immanent account of the motivation for action. Who we are is developed out of the raw material of the world by reflection, what is permitted or enjoined constitutes our future character. The internalism requirement is met by the process of self-conscious reflection itself. We act because of what our ground projects are but our ground projects are made possible by the structure of our self-consciousness. In this sense, our humanity is our capacity to reflect.

Seeing how reflection is the basis of our commitments will now allow us to see a way of turning back Hegel's emptiness charge against the maxim. Because of the structure of our self-consciousness, anything that we reflect on must be seen as fundamentally a self-relation. So, there can be no empty maxim because every maxim that arises for an agent will have the perspective of that individual. The maxim is never empty because it originates in the agent's desires rather than in some abstract impersonal voice and hence is reflected in her to produce a certain outcome, whatever it is. The maxim is thus a situated and contentful thought about a particular situation which is either rejected or accepted. If it is accepted, it is, by this very acceptance, normative for the agent.

Practical Reason and Reflection

Recall that the central criticism that Hegel made of Kant was that the categorical imperative could not move us to action because it was too general. This is a way of saying that Kant's conception of reflection was not normative because it could not move us to specific action and action is always specific, it is

always *this* action. Williams' challenge to Kant brought out the other side of this dilemma, namely that humans *do* have ways of picking out how to act, but that these ways stem from desire, not reason.

Korsgaard's way of addressing these two related issues comes in her theory of action. Before we proceed to examine this in more detail, we should look briefly at the more general way she frames the issue of the rationality of normativity. In *Sources* she argues that the problem of normativity stems from self-consciousness itself. The problem of self-consciousness or reflection is that when we perceive or believe something we can also doubt that what we perceive or believe is correct. In other words, do I really have a *reason* to believe the perception or act on the impulse? The problem of normativity is then, the problem of how to give or find reasons justifying our beliefs or actions.

And this problem, to put it differently, is really the problem of synthesis, of uniting subjective mental states with contingent material or worldly states. This is what Hegel thinks Kant cannot do— provide a set of categories or conceptions for action. But the way Korsgaard puts the problem, adopting Rawls' language of concept and conception, shows that there really cannot be a problem on the most basic level of reflection. Korsgaard writes: "If you recognize the problem [the concept] to be yours, and the solution [the conception] to be the best one, the solution is binding upon you".¹²¹ That is to say that every agent must reflect, mediate the world with her subjectivity and that is what a maxim is. There is

¹²¹ Korsgaard. *The Sources of Normativity*. 114.

(almost) no way of not doing this.¹²² But this, surely, is not what Hegel and Williams meant in their critique of Kant. Their critique was aimed at Kant's claim to provide a 'universal' normative standard according to which each individual could will or act. So the action to be picked out by reflection is not merely any action, but an action which can be acted on by anyone in my situation. Although this question might appear to require a theoretical answer (and this is precisely Hegel's point— how could we ever *know* the depth of any particular situation so that we could be sure it was a good thing to do) Korsgaard insists that universal action must be understood practically. In *Sources*, she thus argues that what distinguishes constructivism from other philosophical perspectives (like realism) is that for constructivism, the answer to an ethical or practical problem in action is not knowledge but more practice.¹²³

Together these two points show us that Korsgaard's notion of reflection in action means that theoretical reason must be subordinated to practical reason. This is not to say that there cannot be purely theoretical knowledge, but it does mean that such knowledge is always at least potentially in the service of practical reason, of action. By arguing in this way, Korsgaard reaffirms what Prauss was deeply skeptical of in the first chapter, namely that practical and theoretical knowledge are in fact one, and that thinking is already acting in practical terms. But this is precisely what I think Korsgaard does want to argue.

¹²² The exception is the case of the wanton, see below.

¹²³ Korsgaard. *The Sources of Normativity*. 36-37.

To sum up, what we are looking for in Korsgaard's account of reflective action is an explanation of how rational reflection (and not some weighing of desires against each other as in Williams) can pick out a (concrete) course of action in the world. The question of just how universal the reflection is (and thus the answer to the richer formulation of Hegel's challenge— the subjectivism charge) will have to be deferred until the next chapter.

The arguments we will examine break into two general parts. First, there is what Korsgaard calls the metaphysical argument which examines the agent from the third person perspective and which is mainly concerned with establishing that, in order to be an agent at all, you have to act on some principle. The argument breaks down into an argument about the impossibility of willing particularly (i. e. being totally determined by desire, which is dealt with in the first sub-heading) and the argument that we do, in fact, will according to the categorical imperative (which is dealt with in the second sub-heading). The second argument, the normative argument, examines what it is like to act from the agent's perspective and shows that in order to act there must be standards of action which are generated by the agent herself. These are the constitutive standards of action.

Acting on Principles

In the *Locke Lectures*, Korsgaard begins by arguing that we are condemned to action. (LL 1 1)¹²⁴ There is no way we *cannot* act, for even not acting is acting. For Korsgaard, the problematic of the primacy of the practical means, quite literally, the problem of action. And here she also distinguishes herself from the beginning of Kant's *Groundwork* by focusing on all action rather than simply on good action. This is significant, since unlike Kant, Korsgaard thinks that all action must on some level be governed by the categorical imperative. The hypothetical imperative is not a separate principle. In order to claim a broader premise than Kant, who sought to show only that morally good action conforms to the categorical imperative, Korsgaard wants only to show that all autonomous action, or what she will call action as opposed to an act, is under the jurisdiction of the categorical imperative.

Korsgaard calls the argument which is meant to establish that whatever we endorse is normative for us and thus constitutes us the metaphysical

¹²⁴ The *Locke Lectures* though (as yet) unpublished offer the best way into Korsgaard's recent work. They partially comprise restatements of arguments given in published material, and I will refer to these when possible. The *Locke Lectures* are available on Korsgaard's web page: <http://www.people.fas.harvard.edu/~korsgaard/#Locke%20Lectures>
I will cite the *Locke Lectures* as LL followed by lecture number and page number.

argument.¹²⁵ This argument is regressive, asking ‘why’ of its premises until an indubitable one is found. Korsgaard summarizes her argument thus:

The necessity of conforming to the principles of practical reason comes down to the necessity of being a unified agent. And if it is correct that agency requires unity, the necessity of being a unified agent comes down to the necessity of being an agent. And if this is correct, the necessity of being an agent comes down to acting. And the necessity of acting is our plight. The principles of practical reason are normative for us, then, simply because we must act.¹²⁶

Let us work forward from the last step and begin with action. What makes action *our* action, that is, what makes the action attributable to us? It must be that we add something to what happens rather than being the passive site of the action. When we are pushed and knock over a glass, we can say that it was not our fault, we did not intend it. But when we pick up a Champagne bottle and pour it over someone’s head, it can plausibly be inferred that we did it. Thus action isn’t merely being pushed around by desires or physical objects, it is contributing to and choosing among, different desires or causes. We are the site of agency only when we act autonomously, when we direct our will.¹²⁷ Having principles according to which we act means having agency, for it means being able to

¹²⁵ It is worth noting Korsgaard’s departure from Rawls here who was uninterested in any sort of metaphysical investigation into what constitutes the self. Korsgaard’s use of metaphysics here is, I think, strongly reminiscent of the transcendental arguments given in German Idealism.

¹²⁶ See LL 1 17. It is unclear to me whether this argument is meant to be a transcendental (and thus synthetic) argument or whether Korsgaard takes this argument to be analytic, the way Kant takes the argument for the rationality of the free will to be.

adjudicate between different possibilities as they present themselves in the material world.

So, if action is only action if it is attributable to us, we must examine what it means to will an action. This question concerns the nature of our ability to will. Korsgaard is looking for the conclusion that willing must occur according to universal principles. To be an agent, you must conceive of yourselves as causally efficacious, that is, you must be the one doing the willing. You are not merely a passive spectator in the battle between desire A and desire B to be turned into action by our will. From this it follows that there must be some principle which allows you to stand against these two desires and choose between them. In choosing which desire to follow, you must regard the decision as expressive of yourself.

So you must identify with one of the two options. But if you will particularly, you cannot say, when asked why you chose one rather than the other, that you did so because you were inclined to do so, since this would mean that you have a principle which enables you to choose. A true particularist would have to embrace the decision in its particularity in a way that is not further describable. Particularist willing thus eradicates the difference between the person and the desire.¹²⁸ And this means that particularist willing, taken as the idea that a person

¹²⁷ See Korsgaard. "Self-Constitution in the Ethics of Plato and Kant" 14.

¹²⁸ Notice that this is not Hume's position since Hume believes that there is an intermediary step between the passions and the external impulse. Reason could not be the slave of the passions if the affectation were immediate. See David Hume. *Treatise on Human Nature*. Bk. 2 Pt. 3 Sec. 3 Para. 5/10.

could will without reference to universal principles and still be an agent (who, by definition, must choose between options), is impossible.¹²⁹ So, in order to will, we need a universal principle of practical willing.¹³⁰

Acting on the Principle of Pure Practical Reason

But if we cannot will particularly, how do we will? We are thus looking for principles according to which we can act. We must will universally, or, at least, provisionally universally, since we cannot know whether all of our premises are true.¹³¹ And, if these principles are to constitute the agent himself, they must be the agent's own. Thus we must ask what kind of law or principle the autonomous agent must give himself.

Korsgaard favors Kant's categorical imperative as a principle. She writes:

The categorical imperative, as represented by the Formula of Universal Law, tells us to act only on a maxim that we can will to be law. And *this*, according to Kant, *just is* the law of a free will. To see why, we need only compare the *problem* faced by the free will with the *content* of the categorical imperative. The problem faced by the free will is this: the free will must have a law, but because the will is free, it must be its own law. And nothing determines what the law must be. *All that it has to be is a law*. Now consider the content of the categorical imperative, as represented by the Formula of Universal Law. The categorical imperative merely tells us to choose a law. Its only constraint on our choice is that it have the form of a

¹²⁹ Korsgaard. "Self-Constitution in the Ethics of Plato and Kant". 26-27. See also LL 2 22-25.

¹³⁰ Korsgaard adds a caveat to the idea of universal willing. We do not, of course, know that our maxims are indeed universal, but we must take them as such. Thus our maxims must be considered provisionally universal. Korsgaard. "Self-Constitution in the Ethics of Plato and Kant". 25.

¹³¹ Korsgaard. "Self-Constitution in the Ethics of Plato and Kant". 25.

law. And nothing determines what that law must be. *All that is has to be is a law*. Kant concludes that the categorical imperative *just is* the law of a free will.¹³²

The categorical imperative thus answers two questions at once; what should I will if I am to remain autonomous, and how shall I will it, that is, what are the constraints on my willing? The important thing to note here, as Korsgaard points out, is that the categorical imperative is not a rule that applies in deliberation. It is deliberation itself and as such it is part of the logic of thought. What thus distinguishes it from the realist approach to moral problems is that it is not to be applied in a theoretical manner.¹³³ The categorical imperative is practical, it is just how we think about things when confronted with a practical problem. The maxim does not give us knowledge, it gives only action.

This point is central to the idea of constructivism in general. The constructive procedure in the metaphysical argument means that the principle of practical reason is interpreted as both form and content of deliberation. Korsgaard argues that practical reason is not an interpretative principle but a constitutive principle. That is, our deliberation furnishes us with a plan of action which constitutes us because action constitutes us. There is no outside to deliberation, no moral truth about the matter, since we just are who we make ourselves to be.

¹³² Korsgaard. "Realism and Constructivism in Moral Philosophy". 114-15. See also *The Sources of Normativity*. 98 and "Kant's Analysis of Obligation".

¹³³ I think the centrality of the constitutive moment of deliberation, though emphasized by Rawls as well, gets lost in his discussion of the Original Position where he writes of the two principles of justice as if they were being applied.

The first part of Korsgaard's argument is thus meant to provide a justification of the principle of practical reason. In order to act, rational beings require principles. It is impossible to act without them and they cannot be particular either. Thus, we have a principle of practical reason. As yet, the principle of practical reason has not been specified. It might still turn out to be either the moral law or some other, instrumental principle. Only the possibility that people are wanton has been eliminated.

Korsgaard's argument for the categorical imperative in the form of the moral law now proceeds by eliminating instrumental reason as well. Korsgaard argues that the hypothetical imperative is subordinate to the categorical imperative much in the same way Rawls argued that the rational was subordinate to the reasonable. Korsgaard calls these two terms efficacy and autonomy.¹³⁴

In order to understand the argument, we must take a closer look at Korsgaard's concept of action. Korsgaard distinguishes between acts and actions. She argues that for Kant, a maxim always includes both the act to be performed and its end. This is so since the question raised by the categorical imperative test is whether there could be a universal policy of pursuing *this sort* of end by *these sorts* of means. Kant is not concerned with the permissibility of act-types or general prescriptions, but with particular acts. For Kant, then, moral worth is ascribed both to end and act together. We can thus distinguish between the act of helping someone and the end of this act, which is the action of doing

something either for pleasure or for duty. Thus, moral worth is ascribed to an act done for its own sake, for the sake of duty, and this Korsgaard defines as an action. It is done because it is good.¹³⁵

Korsgaard distinguishes this view from other contemporary views according to which actions are explained by their purpose. She argues that if actions are really done for their own sake, then the purpose of the action is not always its reason. It may thus appear that every action is done for the same reason: it is worthwhile.¹³⁶ On this view, then, the reason for an action is not external to the action but intrinsic to it. As Korsgaard puts it: “the action is an essentially intelligible object that embodies a reason, the way a sentence is an essentially intelligible object that embodies a thought”.¹³⁷

Here, again, we see the constructivist move according to which the end of the action cannot be external to its purpose. An action, like a deliberation, is what it is by virtue of both its form (its principle, acting from, say, duty or self-love) and its content (the act being performed). To anticipate the argument a little, the action’s content is measured by its own formal standard, whether or not the agent fully succeeds in performing it. In this sense, just as there is no ‘outside’ to deliberation, there is also no ‘outside’ to action.

¹³⁴ LL 3 5.

¹³⁵ LL 1 9-10.

¹³⁶ In this Korsgaard clearly seems to be following Aristotle. See, for instance, McDowell. “The Role of *Eudaimonia* in Aristotle’s Ethics”. 3.

¹³⁷ LL 1 14.

To come back to the distinction between the hypothetical and the categorical imperatives, we can now conclude that an action bound by the hypothetical imperative has its reasons outside itself while the action governed by the categorical imperative has its reasons inside itself. This distinction also explains why Kant thought that the hypothetical imperative was analytic, that is, why he thought that being committed to an end meant being committed to its means. If we determine ourselves to cause an end, we are also committed to cause the intermediate ends (means) necessary to bring the final end about. But proceeding in this way does not tell us anything about the worthiness of the end we are pursuing. The hypothetical imperative is merely effecting an end in the world, any end.

The situation is quite different with the categorical imperative. As we saw, the categorical imperative restricts our acts in certain ways. It requires our acts (from the hypothetical imperative) to be in the service of certain ends. This end is what we have just called being worthwhile. An act is an action if it is restricted to an end that is worthwhile, that is if it is restricted to an internal standard. And, as we saw, this standard must be internal to the action itself, if it is to count as an action, that is, if it is to be autonomous.

And in this way we can see that Korsgaard too 'subjugates' the efficacy of the hypothetical imperative to the autonomy of the categorical imperative. Rawls and Korsgaard both admit that acting on merely the hypothetical imperative is possible, but this kind of act also fails in certain respects (to be examined below). Indeed, Korsgaard goes further than Rawls did by claiming that the hypothetical

imperative is merely a form of the categorical imperative, a law of the will in a causal employment. She concludes this from the fact that the principle of practical reason arrived at above is simply: act in accordance with a maxim you can will as a universal law. This formulation does not make any distinction between the two types of imperatives. This means that the “hypothetical imperative will govern only acts to reach ends, but not the whole package, the act for the sake of the end, if it is worth doing for its own sake”.¹³⁸

This brings the metaphysical argument to a close. We have seen that in order to attribute an action to a rational agent, the agent requires a principle of action which makes the action ‘hers’. This ‘ownership’ of action requires that the agent be unified by a universal principle. Finally, we saw that the principle in question is the categorical imperative, which specifies nothing more than that the law has to have the form of a law and be willed by the agent.

The Normative Argument

We now turn to the normative argument. Whereas the metaphysical argument examined the unity of agency from a formal perspective, the normative argument deals with the perspective of the agent who wills. Korsgaard has earlier said that action is done for its own sake, because it is worth while. The key to

¹³⁸ LL 3 2-3. Korsgaard rules out the prudential imperative as not an imperative at all since, as Kant acknowledges, happiness is not definable. But a principle of maximization requires a theory of what happiness is be in place before maximization can take place and so must be substantive

understanding how the categorical imperative can motivate us is to understand what makes an action worthwhile. In this subsection, we will discuss what makes an action worthwhile in general terms, taking as the model action in the world of things. This argument, however, prepares the way for a development of a conception of standards of actions in the social world as well. This will be the subject of the following chapter. The question is thus not merely the already important issue of how reflection can determine a path of action in the material world, but also of how reflection can determine a path of action in the social world.

Worthiness, Korsgaard argues, is the internal property of the act itself, a set of standards which are autonomous that is, independent of desires and interests. But where do actions get their independence, their ability to judge us back, as it were, to tell us off for not completing them properly? Korsgaard's answer lies in claiming that actions have their own (objective) ends to which we must adopt in performing them. Just as we cannot but understand the world around us except in terms of the categories of the understanding, so we cannot act except under the categories with which actions provide us. If reason does indeed provide us with objective constraints that we must follow, then it can motivate us or constrain us to action.

rather than constructivist. So it is not really a principle of practical reason at all. LL 2 4-7. This is also the central argument of Korsgaard's "The Myth of Egoism".

Self-Constitution, Normativity and Teleology

As mentioned earlier, the notion of self-constitution is an interpretation of the categorical imperative. It is a way of understanding what it means to behave autonomously, and how autonomy is related to normativity. We began with action and concluded that in order for an action to be the work of an agent, it had to be the result of the agent's deliberations. Deliberation, however, requires principles according to which the agent determines herself. This brought us to self-determination and the problem of constitution.

Now, self-determination requires the notion of responsibility since responsibility is an interpretation of the 'ownership' of action. For Kant, what makes a person responsible is that the action springs from his autonomy. I quote Korsgaard:

The exercise of the person's autonomy is what makes the action *his*, and so what makes it an action. And so we get a problem. It is the essential nature of action that it has a certain metaphysical property. But in order to have that metaphysical property it must have a certain normative property. This explains why the action must meet the normative standard: *it just isn't action* if it doesn't.¹³⁹

I interpret Korsgaard as saying that in order for there to be 'ownership' of an action, the action must be both objective (not mine in a certain sense) and subjectively mine (in the sense of coming from me). In order to overcome the apparent gap between my desires on the one hand and objective reality (nature) on the other hand, my action must reflect both its origin in me and its necessary

objective characteristic, the fact that it is intelligible as action for others. Such an action, in other words, requires a synthesis of nature and reason in which the action is both caused by me and conforms to standards other than my subjective nature.

The law which governs the action is called normativity. This normativity is autonomous because the law governing the action comes from the agent's reflection itself. Korsgaard finds support for her interpretation of action as normatively constrained in an objective way in Aristotle's theory of teleology in the *Metaphysics*. For Aristotle, what gives an object its form is its *ergon*, its purpose. To use an example, the purpose of a house is to provide shelter. The house fulfills its function when its parts, the walls and roof, are arranged so that it provides shelter from the elements. Purpose or teleology is what unifies a heap of matter into an object.¹⁴⁰

Objects thus have constitutive standards, standards that are internal. Such standards are not susceptible to challenge from the outside. It *just is* the case that a house must provide shelter. If it does not, it is not a house. Of course it is possible to have a bad house, a house with a leaking roof which nonetheless keeps us mostly dry and shields us from the wind. Depending on the degree of conformity to the internal standard of providing shelter, we can judge whether it is just a bad house or not really a house at all, but a ruin of a house.¹⁴¹

¹³⁹ Korsgaard. "Self-Constitution in the Ethics of Plato and Kant". 14.

¹⁴⁰ LL 1 18-19.

¹⁴¹ LL 1 20-24.

What appears as an internal standard for an object, appears as a constraining rule in the deliberation of a rational beings. In this way, the ‘objective property’ of the object informs the subjective goal-setting of the agent who wants to act. In order to better explicate this, we must revisit the notion of construction. Earlier we saw that constructivism meant a process through which the form and content of a thought constitute one another fully. To illustrate the point, Korsgaard gives the example of pure procedural justice in jurisprudence. She argues that what confers validity on laws is that they have been created by those entitled to create them. Thus the legislature creates laws and they are normative for us because they have been properly created by the authority entrusted with their creation. The agents (legislators) create the laws in the manner in which they are formally constrained to create laws (by voting on proposed bills). This procedure *just is* what makes the law valid.¹⁴² The process of deliberation is itself the synthesis of material circumstances and subjective desires.

In the same way, the outcome of the CI-procedure which really is rational reflection yields a result which must be valid or normative for the agent conceiving it. It has been arrived at in the appropriate manner, by autonomous reflection. The agent wills her action according to the formal constraint of the law she imposes on herself. The reason it does not make sense to question the outcome of the CI-procedure is that it has its authority in the process of reflection the agent engages in— there *can be* no other authority. And because this process is completely internal to our deliberation, Korsgaard says that it is

¹⁴² Korsgaard. “Self-Constitution in the Ethics of Plato and Kant”. 8-10.

thought itself. As the logic of thought, the CI-procedure produces its own authority. There can be no other, if the outcome is to be the agent's own. This is what thinking for yourself is, deliberating according to your own ideas, your own standards.

This constructivist interpretation of autonomy, strong as it is, has now shown that there might be such a thing as a constitutive or objectively rational standard of action. We have now come to the conclusion of the first part of the normative argument: there are rational constraints on the actions of agents and these agents endorse these constraints.

Motivation and Good or Bad Action

So we understand what it is for rational agents to will autonomously— it is to conform to the rational standards given by reason. But there still seem to be many instances in which our actions fall short of this autonomy. This is the problem of *akrasia*, or weakness of will. We may intend to do something and fail. Does this mean that we do not act then according to the categorical imperative, that we are not constrained by the 'objective' standards of action? Korsgaard denies this. She argues that we cannot but be motivated by the objective or constitutive standards of action because that is all we have as a principle of action.

She begins by insisting on the role that reason plays in our autonomous deliberation. Remember that it is not in question whether humans have reason at all (this has already been seen to be the case from the argument against

particularism in which we saw that humans must act under principles), but just whether they are always motivated by reason:

If justice and universalizability are internal standards, then they are not extraneous considerations whose normativity may be doubted. This means that even the most venal and shoddy agent must try to perform a good action, for the simple reason that there is no other way to try to perform an action.¹⁴³

But if we cannot help but try to act according to the laws of reason, does this mean that we must act well or not at all?

In order to answer this question we must understand how the principles of action are normatively constituted. To do this, we must take a short detour to look at Korsgaard's interpretation of Kant's theory of moral motivation, which I also take to be Korsgaard's own. What we are looking for in this account is how the categorical imperative might both motivate us and also fail to motivate us to act in accordance with its constraints. We have seen that, for Kant, there must be some principle beyond our desires and inclinations which determines our will. If desires were immediately motivating for us, we would not be subjects. We saw this in the discussion of particularist willing. Kant believes that only some desires act on us at any one time. He calls these incentives (*Triebfedern*), but they might also be called candidate desires, desires which enter into our consideration for action.

It is these incentives which will form the set of possible actions once one has been endorsed through the process of reflection. That is, we can do them, but we must find out whether it is worthwhile to do them. Here we must note that

reflective endorsement of a desire does not mean that it is actually universalizable, only that it appears to be universalizable to the subject. It is a central feature of Kant's theory of moral motivation that the moral law itself can be an incentive and so motivating.¹⁴⁴ But, by the same token, incentives might be other desires like those of self-love. Whatever the incentive that has been adopted may be, it constitutes the subject normatively.¹⁴⁵

There are thus three types of actions. The first are actions which conform to the categorical imperative. These are actions in the sense discussed above since they are efficacious acts (conforming to the internal standards of the hypothetical imperative; putting one foot in front of the other in order to walk, say) and they are autonomously willed (the maxim is adopted because the end is worthwhile). Second, there are merely efficacious acts. These acts meet their own criteria but are not autonomously willed. Third there are acts which fail to be acts because they do not even satisfy the standards of the hypothetical imperative.

The important point, however, is that we always act in accord with some principle, even if it is not the categorical imperative itself but its derivative, instrumental reason. We are, in this sense, at least minimally rational or reflective. But, Korsgaard wants to insist that all action is done from the same principles, the categorical imperative, and that actions which appear not to be in

¹⁴³ Korsgaard. "Self-Constitution in the Ethics of Plato and Kant". 15.

¹⁴⁴ See Korsgaard. "Kant and Aristotle on Morally Good Action". 206-208. See also LL 3 22.

¹⁴⁵ See Korsgaard. "The Normativity of Instrumental Reason".

accord with the categorical imperative are still actions which sought to conform to the categorical imperatively, but *just were badly done*. I will explain how an action can be badly done in a moment. Let us first look again at why Korsgaard thinks that all actions must conform to the categorical imperative.

Korsgaard's discussion of bad action further buttresses the distinction we saw earlier between acting on the categorical imperative and acting on the moral law. Bad action shows that we can be rational without being moral. For Korsgaard, bad action is not necessarily action which is done for bad reasons such as malevolence, carelessness or self-love. The point of bad action is at once to show that, since action must always be autonomous to be an action at all it must conform to the categorical imperative, but what makes it bad is that it does not conform to the moral law as it would exist in the kingdom of ends. Korsgaard gives the following example: she describes modern constitutional democracy where, through pure procedural justice, unjust laws still come into being. The laws conform to the categorical imperative in the sense that they are autonomously adopted (through the right parliamentary method) but fail in the sense of being unjust or ununiversalizable.¹⁴⁶ Here the deliberation is efficacious without being just. Of course, the badness of the law is only something that becomes apparent after the fact and is then subject to historical revision.

The point is that what makes the action 'bad' is that it has adopted the wrong incentive (content) for action but has willed it in the correct way, by making the outcome of its maxim a law for itself. Bad action does not take reason as its

incentive, but something else. The unity of the subject that is created by the autonomous adoption of the wrong incentive is thus on shaky ground. The agent who has autonomously willed the wrong incentive will find that her actions according to the wrong incentive will come into conflict with those actions which she has willed according to universalizable laws. The bad or defective action is an action which is trying to be unified but just can't quite keep it together. Thus Korsgaard says that a builder building a bad house intends to build as good a house as the good builder, she just cannot quite manage to do so.¹⁴⁷ By the same token, the addict is trying to live the same kind of life as the virtuous person, but just cannot quite manage to do so.

Conclusion

This chapter has dealt with what might be called the 'subjective' side of action, that is, with what makes action possible from the perspective of the agent. We have seen that in order to act, one must be unified under some principle and that this principle must be the categorical imperative. But the question of the 'categorical' part of this imperative is still unclear. It is necessary for the agent to reflect in order to act, that is, it is necessary for the agent to have reason to act. These reasons are furnished by the very process of reflection itself. But the wider question of the intersubjective value of the agent's reasons for acting remains largely unaddressed. We still have not substantively engaged with the objectivity

¹⁴⁶ LL 5 15.

of reflective reasoning. It will be my argument in the next chapter that Korsgaard's notion of the constitutive standards of action together with her argument for a public notion of reason can help us here. That is to say that the problem of egoism is still unresolved. The universal character of reflection, in other words, is still in question.

For now, let me recap the argument in terms of the two criticisms of the Kantian position I mentioned at the outset: Hegel's claim that the maxim cannot pick out a meaningful course of action because it is too abstract and Williams' claim that only desires can motivate. The argument I have presented deals only partially with Hegel's charge since it address only the subjective side of reflection: the possibility of formulating any sort of action through reflection. Korsgaard's vindication of Kant with respect to the synthetic abilities of the maxim clearly centers on the idea that the maxim is not something to be applied to a given situation (an idea which brings with it all sorts of commensurability problems) but rather is the activity of thought itself. That is, if we follow Korsgaard's contention that the solution to the problem of action is the formulation of a plan of action, we can see that there is no lag between the practical problem and a practical solution. Korsgaard thus denies what is, I think, implicitly assumed by the Hegelian critique, namely that the assessment of the problem of action occurs first in theoretical terms and must then be translated

¹⁴⁷ Korsgaard. "Self-Constitution in the Ethics of Plato and Kant". 15.

back into practical terms.¹⁴⁸ While the latter picture entails all the problems Hegel identifies, the former does not, because here theoretical knowledge is subordinated to practical activity. If the maxim really is merely reflection on a practical problem, as has been suggested by Korsgaard, there can be no question that it will yield a concrete course of action.

Korsgaard's insistence on the immanent character of reflection answers not only Hegel's charge against the abstractness of the maxim, but also conforms to Williams' internalism requirement. I interpreted this requirement foremost as a requirement that the authority which turns desires into 'reasons' must be the same authority which experiences the desire or incentive in the first place. Conforming to the internalism requirement in this sense, however, still leaves open the question of what, exactly, 'having reasons' is supposed to mean. It is clear just from the way Williams and Korsgaard formulate what they take to be reflection that they think of the process in different terms. While it is not the task of this dissertation to examine what Williams' means by deliberation in more depth, I think that Korsgaard's references to reflection as originating in the problems of self-consciousness and therefore in Kant's doctrine of the spontaneity of the understanding is clear enough. I will examine how Hegel appropriates this doctrine in the last chapter and I will argue there that Korsgaard's view of reflection is informed by similar considerations. The point for

¹⁴⁸ This conception also informs Prauss' doubt that practical and theoretical reason could really be one and the same thing. See chapter one.

now is that Korsgaard's view of reflection is immanent and that she sees reason as self-authorizing.

It has also become clear in this chapter that Korsgaard forcefully rejects the idea that desires can motivate directly. I do not think Williams' claim that commitments motivate should be taken as endorsing precisely the sort of direct motivation which has here been shown to be impossible. I thus propose to read Williams' insistence on the motivational power of personal commitments as a moral claim, or rather, as an ethical claim made against the Kantian claim of morality. We should read Williams' charge as centering on the question of whether our commitments to our projects can be shared, rather than saying anything about the possibility of reflection itself.

CHAPTER V

INTERSUBJECTIVITY IN KORSGAARD

In this chapter I would like to take up the problem we saw articulated earlier by both Williams and Hegel, namely the question of the scope of normativity. The theory of reflection offered by Korsgaard in the previous chapter showed that normativity exists, that is, that we do need reasons for action. That discussion, however, left open whose reasons count for me. For, as I have noted repeatedly, Williams insists that the only types of reasons that are truly normative for me are my own reasons. In this chapter I will take up the charge of subjective normativity or of egoism, as it is usually formulated in the Anglo-American literature. I would like to present this challenge in the form of Hegel's theory of subjective action because Hegel presents both a fresh formulation of the problem of egoism (which, however, can be assimilated to Williams' concerns) and also takes the problem to its dialectical limit, which will allow us to see what is needed to avoid this challenge. (This will be the subject of the next chapter.)

The second half of the chapter will be devoted to Korsgaard's arguments against subjective normativity or egoism. These arguments seek to show that egoism, conceived of as having private reasons, does not yield a stable conception of agency. Only an agent who is able to take into account other people's desires as his own is able to act on the categorical imperative and only this will make him a coherent agent. Finally, I will evaluate how effective

Korsgaard's defense of the publicity of reasons is in turning back Williams' Gauguin example in "Moral Luck".

Hegel's Account of Action

I will now give a brief reconstruction of Hegel's theory of subjective action, culminating in his criticism of conscience. Hegel's account is similar to Williams' in that it accepts the internalism requirement and insists that the articulation of a subjective ground project is of constitutive significance for the agent. The radical internality of his ground project is, as in Williams, given a privileged space in Hegel's interpretation, and constitutes pure subjectivity. But by formulating the subject's interest thus, Hegel also points to the limits of such internality. The limits of the subjective norms Hegel diagnoses are similar to Williams' fatalistic claim that the happiness of an individual really just depends on whether his ground project happens to line up with currently acceptable norms. Both are forms of the impasse of subjectivity.

Michael Quante has distinguished three elements in Hegel's theory of action: (1) the subject's perspective on the action as being done for the subject itself, (2) the normative nature of the action, its relation to the principle according to which it is carried out (that is, whether it conforms to this principle) and, (3) the demand that the action be recognized by others as being good.¹⁴⁹

¹⁴⁹ Quante. *Hegels Begriff der Handlung*. 128-30.

These three stages in Hegel's theory correspond roughly to the development of the subject from reflective endorsement to social recognition. That means that we are moving from the subject's interaction with the world of things (which includes treating other people as 'things' as well) to an interaction between subjects which is characterized by recognition or misrecognition.

Hegel begins the Morality section of his *Philosophy of Right* by claiming that in order for the subject truly to be a subject, it must determine itself according to its own standards, rather than according to the standards of others.¹⁵⁰ Hegel calls this the right of the subjective will: "The will which is determined as subjective and free for itself, though initially only concept, itself has *existence* in order to become *Idea*. The moral point of view therefore takes the shape of the *right of the subjective will*."¹⁵¹ The right of the subjective will means a retreat of the subject into its own willing and away from the influences of others. This marks both a rejection of the (arbitrary) law that others (perhaps the Hobbesian state) wish to impose on it, and a rejection of the subject's merely physical needs and inclinations.

¹⁵⁰ GPR §106. The "Morality" chapter is the second section of the *Philosophy of Right* and deals with Kant and the critique of moral subjectivism more generally.

¹⁵¹ "Der als subjektiv bestimmte, für sich freie Wille, zunächst als Begriff, hat, um als Idee zu sein, selbst Dasein. Der moralische Standpunkt ist daher in seiner Gestalt das Recht des subjektiven Willens." GPR §107. See also GPR §117 where Hegel explains the right of the will as the right only to be held responsible by others for what the will could have known its actions would lead to.

The right of the subject, however, at least as initially conceived, is empty, because there is nothing binding the application of these rights. The subject's will just *is* its action. But as we examine the rationality of the subject we are also looking for ways in which the subject limits itself by its own commitment. The right of the subject must therefore be something like a duty of the subject. Hegel thus construes the internalism requirement as duty: an act is a duty when it is normative for us. Hegel's point, like Korsgaard's, is that having a principle is not a static relation to the world but rather an active engagement with it. Normativity arises when the subject imposes certain limitations on its actions, i.e. acts autonomously. These limitations, Hegel contends, are the result of the subject's interaction with the world.

The requirement of the will to assert itself in nature, to transform nature according to its principles, is experienced by the will as an imperative. In acting autonomously, the will must will itself. But because the subject is also part nature, it must appropriate the fact of law for its own purposes and claim the lawfulness of nature as its own. This is what it means for the subject to determine itself in accordance with the representation of laws.¹⁵² And this representation of laws is experienced as normative or as an 'ought'. I must take the means to my

¹⁵² Kant. "Everything in nature works in accordance with laws. Only a rational being has the capacity to act *in accordance with the representation* of laws, that is, in accordance with principles, or has a *will*." "Ein jedes Ding der Natur wirkt nach Gesetzen. Nur ein vernünftiges Wesen hat das Vermögen, *nach der Vorstellung* der Gesetze, d.i. nach Prinzipien, zu handeln, oder einen *Willen*." GMS 412.

desired ends, for desire itself is only properly constituted within the bounds of the resistance I experience from the world.

Lastly, the question of the good arises. The good, for Hegel, is a more social version of the well-being each agent desires for himself. The good is what is desirable for all. It is part of Hegel's argument that there is no categorical shift from desire to the ethical relation since there is no fundamental difference between the world and the other subjects in it. What is desirable for all, however, is constituted not in an a priori way but is developed out of the interaction between subject and world, now conceived of as both the material world and the social world. This is where the impasse of subjectivity occurs: the individual's desires might not be reconcilable to the demands of society.

The Subjectivism Charge: Conscience and Gauguin's Ground Project

This, I think, is where we might locate Williams' Gauguin. The claim that Williams makes for Gauguin's need to act on his ground project appears to be something like the following. To quote Hegel: "For the *subjective* will, the good is likewise absolutely essential, and the subjective will has worth and dignity only in so far as its insight and intention are in conformity with the good."¹⁵³ And now to paraphrase: the subjective will's conception of its own good is so essential to it that the subjective will's value consists entirely in seeking to realize its conception of the good. We might call this the Luther moment. ("Hier steh' ich,

ich kann nicht anders".) The point though is, I think, clear enough. Gauguin's conception of the good, his ground project, is so important to him that he is prepared to break with the conventions of civil society and to leave his wife and children in order to seek to fulfill his own-most commitments.

Hegel connects this Luther moment with conscience in a deprecating way which fits rather well with Williams' own skepticism about the ability of society to furnish us with appropriate moral guidelines. Hegel argues that conscience is the ability to determine the good for oneself in all vehemence and conviction, but that this determination is completely subjective. This way of putting it pits Gauguin's subjectivity against the objectivity of social mores in just the way Williams intends it; the idea of commitment to one's ground project is overriding in the same way we speak of the normative power of conscience, which plagues us even after we have gotten away with breaking some rule. But here it is the breaking of a social norm that has become central to the identity of the agent. Conscience is thus fundamentally an anti-moral attitude in both Williams and Hegel. And it can (and, according to Hegel, will) become immoral because, in taking itself to be *moral*, it goes against the ethical, which is always conceived of as objective.

A further element that Hegel's analysis of the subjective will shares with Williams' Gauguin, is the idea that the subjective will can and must be judged by those around him. Hegel writes: "The right to recognize nothing that I do not perceive as rational is the highest right of the subject, but by virtue of its

¹⁵³ "Für den subjektiven Willen ist das Gute ebenso das schlechthin Wesentliche, und er hat nur Wert und Würde, insofern er in seiner Einsicht und Absicht demselben gemäß ist." GPR §131.

subjective determination, it is at the same time *formal*; on the one hand, *the right of the rational*— as objective— over the subject remains firmly established.”¹⁵⁴

Hegel is thus pitting the right of the subjective will against the rational or objective order of things. Though Williams would clearly not agree that there is a rational order of things that Gauguin violates, it does seem that he agrees that Gauguin is going against the (somehow arbitrary) norms of society. Indeed, Williams explicitly denies the idea that Gauguin is amoral.¹⁵⁵

But to say that Hegel is contrasting the rational order of the world with the subject’s particular notion of the good is to overstate the point. For what is precisely at issue here is whether there is a rational order of the world which the subject can seek guidance from or whether all we have to go on is our own commitments. This is clearly where Williams and Hegel (and Korsgaard) part company— for while Williams seems to think that it really *is* a matter of luck what sort of subjective commitments we come to have, Hegel thinks that the sorts of commitments we have are a function of historical development. For Williams luck seems to be something truly arbitrary while for Hegel it seems to be something historically determined.

Lastly I want to touch on an issue which is present as a possibility in Williams and which is brought out explicitly by Hegel, namely that the right of the subjective will really is evil and not good at all. Williams acknowledges this point

¹⁵⁴ “Das Recht, nichts anzuerkennen, was Ich nicht als vernünftig einsehe, ist das höchste Recht des Subjekts, aber durch seine subjektive Bestimmung zugleich formell, und das Recht des Vernünftigen als des Objektiven an das Subjekt bleibt dagegen feststehen.” GPR §132. Anm.

when he says that Gauguin's ground project may never be justifiable, even to Gauguin himself. Here we should not conceive of evil as intentional evil but rather circumstantial evil, brought about by the misguided actions of different actors. Hegel writes:

Where all previously valid determinations have vanished and the will is in a state of pure inwardness, the self-consciousness is capable of making into its principle either *the universal in and for itself*, or the *arbitrariness* of its *own particularity*, giving the latter precedence over the universal and realizing it through its actions— i.e. it is capable of being *evil*.¹⁵⁶

The problem with acting on one's subjective commitments means that one might be acting contrary to the interests not only of society in general, but also of oneself. Personal commitments are here seen as trumping social and personal norms.

Hegel locates the origin of these personal commitments in nature itself, in the irreducible specificity of the drives which, as he notes, can be construed as either *good or bad*— and for Hegel good or bad are relative to the level of historical social development. This is indeed how Williams conceives of them. It is a matter of luck to what extent your personal commitments or drives coincide with the general conception of what is acceptable in a given society.¹⁵⁷

¹⁵⁵ Williams. "Moral Luck". 23.

¹⁵⁶ "Das Selbstbewusstsein in der Eitelkeit aller sonst geltenden Bestimmungen und in der reinen Innerlichkeit des Willens ist ebensowohl die Möglichkeit, das an und für sich Allgemeine, als die Willkür, die eigene Besonderheit über das Allgemeine zum Prinzip zu machen und sie durch Handeln zu realisieren— böse zu sein." GPR §139.

¹⁵⁷ Nussbaum argues that Williams' target in his critique of morality is the Victorian morality system which was particularly repressive in sexual matters. See Nussbaum. "Bernard Williams".

Compounding the importance of this point, both Williams and Hegel take the expression of conscience or one's ground project as a fundamental expression of one's freedom.¹⁵⁸ The more pessimistic dialectical point which Hegel draws from the notion of possibility or luck is that the will, which might be at odds with society, *will* come to be at odds with society at some point. This, however, means that such a will *must* be seen as evil.¹⁵⁹ I think this pessimistic point is echoed in Williams' desire to roll back the brutal hegemony of an entrenched morality system to allow more space for individual freedoms.¹⁶⁰

Hegel's strong formulation of the problem of conscience and Williams' equally strong formulation of the need to fulfill one's ground project seem to suggest that we have come to an impasse in terms of articulating a theory of normativity which is binding to all the way it is binding to each. This does not preclude a political solution like the one Rawls suggests through the overlapping consensus in *Political Liberalism*, but it does cast serious doubts on the idea that commitments must be shared to be normative. But the political solution at least allows the possibility of a gradual convergence of different types of personal commitments, thus reducing the relative 'evil' of each.

¹⁵⁸ Hegel writes that the origin of evil lies "in the speculative aspect of freedom, in the necessity with which it emerges from the *natural* phase of the will and adopts a character of *inwardness* in relation to it." "in dem Spekultativen der Freiheit, ihrer Notwendigkeit, aus der *Natürlichkeit* des Willens herauszugehen und gegen sie *innerlich* zu sein". GPR §139 Anm.

¹⁵⁹ GPR §139. Anm.

¹⁶⁰ Williams. *Ethics and the Limits of Philosophy*. Ch. 10.

I would like to devote the rest of the chapter to an examination of how Korsgaard responds to the challenge articulated above. A key to Korsgaard's response is to see that she moves the standards of action from merely technical standards to intersubjective standards. That is, she thinks that the way we interact with others has its own set of standards and that these standards come from conforming our actions to the categorical imperative. By taking other people's reasons into account, we treat them as ends rather than as means. It is, however, a matter to be investigated whether the *possibility* of taking others to be ends in themselves really helps us with the problem of morality as I have formulated it.

The Intersubjectivity of Reason in Korsgaard: The Wittgensteinian Argument

In *Sources* Korsgaard presents an argument, itself a revision of an earlier argument, which seeks to refute the charge we just saw in Williams and Hegel, and to establish that the normativity of reason obtains not just for the self but for all others, at least in principle.¹⁶¹ Though this argument has been criticized as too thin to provide the solution to the problem of intersubjectivity itself, it is an important step toward the more powerful argument for the intersubjectivity of reason given in the *Locke Lectures*. The argument in *Sources* also appropriately formulates the problem of autonomy and its necessary relational character. As such, it shares many of the features of Hegel's view of intersubjectivity itself.

¹⁶¹ See. Korsgaard. "Reasons We Can Share".

Williams' claim that reasons for action are essentially personal or private was formulated in *Ethics and the Limits of Philosophy* as the problem that, though the fact that you have a plan gives you reason to pursue your plan, and the fact that I have a plan gives me reasons to pursue my plans, this does not mean that I am bound to consider *your* plans a reason to change *my* behavior. Williams even suggests that it might be the case that if I were you, I would *want* to interfere with my goals (because, presumably, my activities limit your goals in some way).¹⁶²

The problem with this way of formulating the objection, Korsgaard notes, is that the objection itself relies on a premise that is widely assumed in Anglo-American philosophy, but not itself argued for, namely that reasons are fundamentally private and must be shown to be public in order to establish morality.¹⁶³ By seeking to refute this argument and to establish the intersubjectivity of reason, Korsgaard is also taking issue with Williams' formulation of the internalism constraint according to which only desires motivate our actions, a position which itself implies that desires are private. I think we should thus take Korsgaard's argument against private reasons as a further argument for the necessity of conceiving of reason-giving according to her model of reflection rather than according to the sub-Humean model Williams employs.

¹⁶² See Williams. *Ethics and the Limits of Philosophy*. 61.

¹⁶³ Korsgaard. *The Sources of Normativity*. 132.

Korsgaard's argument is borrowed from Wittgenstein and relies on a parallel between language and thought.¹⁶⁴ Korsgaard interprets Wittgenstein's private language argument as stating that language is inherently relational because it is normative.¹⁶⁵ To say that X means Y is to say that one *ought to* take X to be Y. That is, having a language means having a certain type of normative relation to oneself, namely that of giving oneself instructions or rules on how a certain term should be used. To use a term correctly thus means to use it according to a specific rule which one is not free to change.

Reasons are relational in just the same way. To have a reason is to determine oneself in a certain way. The publicity of reason comes in just where it does in the private language argument. Though it is perhaps most common only to articulate one's reasons to oneself just the way one only explains one's use of words to others when pressed, reasons and words have an intrinsically sharable quality because they must provide a law according to which we bind ourselves.

The argument also models the intersubjectivity of autonomy in the sense that giving oneself a law means treating oneself as someone else, someone who is capable of being bound by someone else's commitments. Korsgaard here cites Derek Parfit's notion of the serial character of person whose relation to himself is not so different than the present self's relation to other current selves.¹⁶⁶ That is,

¹⁶⁴ Though Nagel raises the objection that egoism does not violate publicity, I think that Korsgaard's response is adequate. See *Sources of Normativity*. 208 and 223-228.

¹⁶⁵ The argument is made in *The Sources of Normativity*. 136-39.

¹⁶⁶ See section III of Derek Parfit. *Reasons and Persons*. It is telling that Williams' argument in "Persons, Character and Morality" relies on Parfit's notion of identity as well.

to commit myself to something I say means also that I can commit myself to something someone else does.

But the Wittgensteinian argument establishes only what we already assumed to be the case in Korsgaard's argument about the *possibility* of reflection: that it is at least in principle possible to give reasons to others. So it now remains for us to see why we *should* seek to take other peoples' reasons into account. This piece of the puzzle comes in the form of the argument for self-constitution, and internal and external justice in the *Locke Lectures*. In this argument, the notion of personal integrity will play a central role.

The Intersubjectivity of Reason in Korsgaard: Taking Others' Reasons as One's Own

In *Sources* Korsgaard used Wittgenstein's private language argument to show that reasons cannot be private because in order to be normative, they have to be at least capable of being made public, or of being articulated. The last installment of the *Locke Lectures* seeks to establish the same conclusion but gives a more detailed account of interaction than was available in *Sources*. The new argument is meant to pick up on the reflection argument we have already examined by showing that justice is really a further demand of the unity of agency— and this has already been established in the normative and metaphysical arguments. I take this to be an expansion of the previous argument for the possibility of interaction in the sense that the new argument states that because we *should* seek to be unified agents, we *should* seek to act according to the categorical imperative.

I would like to present the argument in a different manner than Korsgaard does in her lectures. While Korsgaard seeks to show that taking others' reasons as motivating reasons is a consequence of the need to be consistent in the way one regards one's own goals and those of others, I would like to frame the argument in terms of the internalism requirement. Putting the argument that way will help us see how Williams' critique of the Kantian position is both accommodated and refuted by Korsgaard.

Let us begin with reflection. As we saw earlier, it is the nature of self-consciousness to reflect on the perceptions and desires which arise both within the subject and outside of it. Choosing which action or belief to pursue means that the subject must endorse a particular desire or perception over others. This endorsement gives the agent unity and constitutes the agent's identity. Reflection, to put it simply, gives us reason to act. We also saw that Williams' critique of the Kantian position implied that deliberation entailed no such reflective unification of the subject but rather consisted in the ordering of the subset of inclinations and desires in a certain way. Korsgaard's argument for intersubjectivity or the ability for people to take other people's desires as motivating relies on the notion of reflection. I will first lay out Korsgaard's argument for intersubjectivity and then (in the next section) proceed to discuss how Williams' notion of deliberation does not yield intersubjectivity.

Korsgaard's argument simply appeals to the fact that the subject experiences desires as external. If desires and inclinations must be endorsed by the subject in order to become reasons for action, and all desires are in some

sense external to reflection, then it does not really matter where the desires come from. They might stem from my own intentions and desires, but they might also stem from the wants and reflections of others. Korsgaard writes:

I must interact with the conscious inhabitants of my body, because I must act with my body. But I may also interact with other people, and when I do then their reasons, as well as my own, become as it were incentives in the deliberative process that we undertake together, resources for the construction of our shared reasons. So taken in that way, the category of my reasons doesn't exclude an identification between my reasons and the reasons of others.¹⁶⁷

What is important is that all desires must be endorsed by me. There is then, no reason why your request that I pick you up from the airport should not trump my desire to sleep in. Both candidate desires have to be reflectively endorsed. Where the desire comes from makes no difference.

This argument is meant to refute the claim, put slightly differently than before, that egoism is capable of being universalized.¹⁶⁸ That is, it is possible for me to acknowledge that your projects are normative for you all the while denying that they are normative for me. We examined this claim above in terms of the idea that reasons were private. But if Korsgaard's take on the internalism requirement is correct, then it is at least possible that reasons are public, that is, that your interests or reasons should be acknowledged to be reasons for me.

¹⁶⁷ Korsgaard. LL 6 21.

¹⁶⁸ On this, see Nagel's contribution to *The Sources of Normativity*. Lecture 6.

Korsgaard gives a useful example. It comes in two varieties, one assuming the publicity of reasons, the other assuming their privacy. Let us look at the publicity of reasons first. A student and a teacher want to meet.¹⁶⁹ This creates a practical problem to which meeting would be the solution. If the teacher suggests a time which the student cannot make, and the teacher genuinely seeks a solution to the problem common to both, the teacher will take the student's inability to make the suggested time as a *reason for her* to suggest a different time. This is because the meeting is a problem common to both, a problem which can also only be solved if both make the meeting. In seeking a solution, both student and teacher will take each other's scheduling problems as their own problems.

In deliberating in this way the student and the teacher's wills will be fused, not in some metaphysical way, but by each taking the other's reasons to be her own. Korsgaard writes:

When we interact with each other what we do is deliberate together, to arrive at a shared decision. Since the conclusion of a practical syllogism is an action, the result is an action that we perform together, governed by a law we freely choose together. The free choice of this law is an act that constitutes our unified will and makes shared action possible.¹⁷⁰

In taking the student's request to meet seriously, the teacher acknowledges the need for a meeting to be her own. That is what it means for an

¹⁶⁹ Korsgaard. LL 6 17-19.

¹⁷⁰ Korsgaard. LL 6 15.

agent to take another agent's needs or desires seriously— it means making them her own. Consequently, the problem can only be solved if both can make the meeting.

From this argument Korsgaard concludes that we *can* treat people as ends in themselves for that is what we are doing when we take others' reasons as our own. This is not to say that we always do so. This part of the argument rearticulates the Wittgensteinian point about the possibility of intersubjectivity.

Deliberation as Action

The above example leaves us wondering whether it is more than mere chance whether we interact with others in the way Korsgaard says is possible. A way of addressing, if not answering, this problem presents itself in the theory of the constitutive standards of action. Though Korsgaard does not make this explicit, I take it that the argument about the publicity or privacy of reasons is analogous to the argument about good and bad actions. In discussing the constitutive standards of action, Korsgaard argued that there was really no such thing as intending to build a bad house. When one acts, one acts on something that is worthwhile and what is worthwhile is building a (good) house. You build a bad house only if you don't know how to do it right, but you cannot intend to build a bad house (though you can intend to build a house that will stand just until the house is sold, but in that case, your intention is fraud and not building a house).

The case of deliberation is much the same. Let us look at the unsuccessful example of deliberation. The version in which there is not public

deliberation is the one we might be more used to imagining when there is a differential of powers as commonly exists between teacher and student. This scenario takes the same starting point, that of a student seeking a meeting with a teacher. But this time, the teacher does not take the student's reasons as her own. She does not really care whether she meets the student or not. The request for a meeting still poses a practical problem but the problem just is not the same for both student and teacher. The teacher just wants to get on with her work and does not care whether she meets the student or not. What then passes for interaction or deliberation is nothing of the kind, but is rather two individuals reflecting on their own desires in the same space and hoping for the coincidence of agreement. Each treats the other's reasons not as reasons but as obstacles to what he or she wants to accomplish. It is important to note that though the student might still believe that he is trying to solve the problem together with the teacher, the teacher's private goal renders the student's goal private as well, because the teacher's obstinacy will mean that the student experiences the teacher's reasons as obstacles as well.

What is going on here is that the student and the teacher are not pursuing the same thing and are thus not cooperating on the same project. This could be because the teacher does not think she should be burdened with having to meet students and is thus actively trying to avoid meeting the student. This case is analogous to the fraud case above, because the teacher is simply pursuing a different goal. (Here it does look like the teacher has private reasons or, at least, reasons she does not want to share with the student— we will return to this

below.) Or the teacher just might be bad at social interaction, in which case the teacher *does* intend to meet the student and acknowledges the force the practical problem of meeting has on her, but she just can't quite make it work. The failure of the meeting is unintentional, just the way building a bad house is unintentional.

But there is also a difference between building a house and deliberating with someone else. This difference can best be captured in terms of the difference between the hypothetical imperative and the categorical imperative. Building a house is an action which must conform to the technical imperative which is the hypothetical imperative. Though there might be certain intersubjective standards involved in this, it is clear that the technical imperative is overriding. But deliberating with someone else is only a technical matter to a limited extent: it cannot be done on its own and it has no particular procedure. I cannot claim that I was deliberating with you but you just weren't listening. It just isn't deliberation unless both people are involved and know themselves to be so.

If the analogy between action and intersubjective deliberation holds up, then it looks like there really is no way in which we could not seek to deliberate, to take other people's reasons into account when we reflect. The example of private reason is then not really an example of reason being private, but of reason being insufficiently attentive to its public function. Because of this, I want to interpret Korsgaard's claim that deliberation is hard work as a claim about the relative difficulty of reaching a solution to a practical problem rather than a claim

that it is hard to overcome one's egoism and step into intersubjectivity.¹⁷¹

Korsgaard's claim is that if we deliberate reflectively then we also deliberate publicly, because the process of reasoning is open to all possible desires. So there is no way of ever really having private reasons. This is a strong claim and it appears to contradict the case of the teacher who did not want to meet with the student in order to have more time for her own research. But it does not.

In *Sources*, Korsgaard has argued, convincingly I think, that merely by being in a room with you I obligate you somehow.

If I call out your name, I make you stop in your tracks. (If you love me, I make you come running.) Now you cannot proceed as you did before. Oh, you can proceed, all right, but not just as you did before. For now if you walk on, you will be ignoring me and slighting me. It will probably be difficult for you, and you will have to muster a certain active resistance, a sense of rebellion. But why should you have to rebel against me? It is because I am a law to you.¹⁷²

From this I think we should take it as having been established that on the internalist model of reflection Korsgaard offers, intersubjectivity is possible and even common. This does not, however, mean that there cannot be cases like that of Gauguin in which one puts what appear to be purely subjective desires ahead of all the reasons supplied by others—Gauguin's leaving despite his family's pleas for him to stay and support them.

¹⁷¹ Korsgaard. LL. 6 19.

¹⁷² Korsgaard. *The Sources of Normativity*. 140.

The Divided Self

I would now like to explore why Korsgaard thinks there can be no such thing as private reasons in connection with Williams' notion of the ground project. Doing so will provide an additional argument supporting Korsgaard's claim that all reasons must be public and also provide a critique of Williams' claim that ethical commitment is subjective. However, in taking up the problem of the seeming-egoist we should note that we are moving into a slightly different complex of problems than before, for we are now dealing with someone who turns his back on the reasons of others though, as we have just established, he is very well able to hear and appreciate them. I want to suggest that we are now moving away from the theory of what it means to be a subject capable of taking up the reasons of others, and toward a theory of how desires are furnished by the world around us for *actual* endorsement. In doing so, we are also moving from the question of possible justice to the question of real or concrete justice. This move will also make it necessary to say something about how political organization is different from moral capacity. (Recall also that one of the themes I have been pressing is the need to abolish a moral/non-moral partition of the world.)

Korsgaard begins discussing the problem of private reason by introducing the model of a city, a model shared by Plato and Kant. She argues that a city which is just to its citizens cannot very well be unjust to others, or, if it is, it will not last very long. Here having private reasons means being unjust outwardly, as

in the student-teacher example above.¹⁷³ In both cases it seems to be possible to act on one's own reasons against those of others, but both sets of actions undermine the social network upon which those very actions depend.

An example taken from utilitarianism illustrates the problem. As a utilitarian you might believe that it is good for people to believe certain things and you may try to get them to think these things. It might occur to you that you too should believe what it is useful to believe. But you can't believe what you don't think is true. But if you find that it is not useful for you to believe the truth, you cannot comply with this and your inward justice or self-respect and consequentially your outward justice or respect for others is in disarray. The problem stems from the fact that your reasons for having people believe certain things will not pass the test of publicity. This is just why the teacher in the second example must keep silent her true aim of not meeting with the student. Korsgaard argues that the problem with the utilitarian view is that the standard of reflection articulated earlier holds you to a more stringent standard of reason-giving than the utilitarian model allows; while the utilitarian model seems to endorse self-deceit, Korsgaard's reflective model shows that, though we might condone deceit, we cannot self-consciously engage in self-deceit.¹⁷⁴ This is so because in reflection all desires are equal until they are endorsed by reflection and constitute the subject.

¹⁷³ Korsgaard has in mind Plato's *Republic* and Kant's "Perpetual Peace".

¹⁷⁴ Korsgaard. LL 6 5-6. See Williams. *Ethics and the Limits of Philosophy*. Chapter 6 for a similar critique. See also Kant's discussion of the realist's position in "Perpetual Peace".

So, what would self-deceit mean? On Korsgaard's view, self-deceit is an oxymoron because it is only through endorsing certain desires over others that the self comes into existence.¹⁷⁵ Someone who deceived him or herself would be a bundle of desires without a self or a self divided. But this, Korsgaard argues, is precisely what the egoist is. To illustrate the point, Korsgaard takes the example of the Russian nobleman, taken from Parfit, that we have already seen Williams use in "Character, Person and Morality".¹⁷⁶ Recall the scenario: the idealistic nobleman who will come into an inheritance when his father dies wants to give his estate to the peasants. Anticipating, however, that, as he gets older, he will become more conservative, he arranges for his estate to be disposed of in accordance with the wishes of himself and his wife and instructs his wife to remain true to his present self by insisting against his future self that the estate be given to the peasants.

My analysis of this example does not precisely follow Korsgaard's but I think that the point can be gotten at in my way as well, with the added result that we will see the parallel to Williams' Gauguin example. The problem with the nobleman is clearly that he does not know what he wants. On the one hand he wants to give his future estate to the peasants, on the other hand, he anticipates wanting to keep it. So he cannot make up his mind. The solution of asking his wife to give the estate to the peasants despite his (future) self's wishes is only an

¹⁷⁵ Here we are not talking of psychological or unconscious states, just of states which appear in reflection.

¹⁷⁶ See Parfit. "Later Selves and Moral Principles".

apparent solution, for he will also be asking her (in the future) to keep the estate. The two requests, though temporally set apart, have the effect of canceling each other out.

The problem here is similar to the problem we saw in chapter four when we examined Korsgaard's particularist willing argument. There the argument was that if you are made up only of competing desires, there will be no *you* which can act. Here the case is the same: the nobleman cannot act on either desire because he is unable to make a law to which he is willing to bind himself. Asking his wife to be a law for him would only work if he could already be a law for himself. The problem faced by the nobleman is thus that his self lacks the authority to give itself laws. The nobleman cannot do anything because he cannot constitute himself through action. In this sense, the divided self is not a self at all. The point to be understood here is that, given the reflective endorsement argument, we have no choice but to interpret the nobleman's or Gauguin's predicament in social or intersubjective terms.

Constitutive Standards of Ethical Action

So to say that the nobleman is not a self is to put the problem a little too strongly. Surely the nobleman has desires on which he does act. But it is his overarching desire, his desire to do good for both himself and for the peasants, that puts him in a moral quandary. So the problem the nobleman faces is not the problem of reflection in general, but of unifying his intentions and subjective desires with those of other subjects around him. He experiences the needs of the

peasants as reasons to give away his estate just as he experiences the demand of his future (perhaps ailing) self as reasons to keep it for himself. Korsgaard's argument about the divided self is thus an argument about the difficulty the agent experiences in trying to satisfy competing demands, which *are* reasons for him, in a situation in which only one of the demands can be met. The nobleman's problem is thus that his two ground projects threaten to split him in two. He is torn to the extent that there threatens to be no self left to adjudicate between projects. And this problem, I want to argue, stems from the organizational structure of the social world itself.

This is where the notion of moral luck comes in again. For it is a question of luck which moral dilemmas we are faced with, which competing reasons we have to judge between. From this perspective, I think we can interpret Williams' point about moral luck as a question about the origins of what we happen to value, what is available to us as reasons for action. Apart from the clearly biological needs of nutrition and shelter, the great majority of reasons for actions are cultural or social.

Given the political nature of the problems faced by the nobleman, Gauguin and even the teacher, we should now look deeper into how the social world is constructed such that different ground projects could come into conflict with each other. What I ultimately want to argue is that the historical nature of the standards for action or practices are not wholly cohesive, which means that different

practices will conflict with each other. This is where we should locate both the possibilities of moral luck and historical progress.

There is a continuum between what we might term technical acts like walking and ethical actions like deliberation. All but the most basic technical (biological) actions are historically determined.¹⁷⁷ The more complex the action is, the more socially determined it will be. Thus, the practice of building a house has changed in the past two hundred years and a house built two hundred years ago would now no longer pass a building inspection. This means, to put it in terms of reflection, that the desires one must endorse in order to accomplish the action of building a (good) house are now different than they were two hundred years ago. This has both to do with technical standards and with more general practices of modern life, like where the house should be built, how it is decorated, etc.

Though on a continuum with the more technical imperatives of building a house, what counts as interaction today has also considerably changed from what it was 200 years ago. We talk to different people than we might once have and we speak to them with more respect than we might once have. (The need to put the example whiggishly is, I think, one of the consequences of Korsgaard's and Hegel's views. We will come back to this.)

To understand the political significance of social institutions or practices, we must examine the idea of reasons for actions from the other side. For it is not

¹⁷⁷ Though the Monty Python sketch about the ministry of funny walks does its best to undo even the relatively fixed notion of walking.

only the case that in order to accomplish our desires we must endorse them in certain ways, do X for the sake of Y. It is also the case that the practice of Y tells us that X is the appropriate way to get Y done. This means that the institutions that exist determine which means should be endorsed to accomplish which ends. To give an example, the end of maintaining discipline in the classroom gave teachers reasons to use the paddle. Today, the same end gives us reasons to try to reason with students. (Again a whiggish story.) So, I think on this reading we would want to say that the end of satisfying his ground project of artistic self-expression gave Gauguin reasons to leave his family and move to Tahiti. The ground project of living with his family, we might add, also gives Gauguin reason for staying. The problem, then, is that Gauguin's different ends are at odds with each other in such a radical way that pursuing one means completely relinquishing the other. And this, I have been arguing, is a political problem. It is a problem to be solved collectively.

What still remains unclear in Korsgaard's account is how we come to have certain reasons and how, concretely, we are pushed toward resolving conflicting moral claims using them. The standard of wanting to have a whole self is insufficient on this count.

Conclusion: The Question of Justice

In this chapter I have presented Korsgaard's theory of intersubjectivity as the response to the charge that reasons are normative only for the agent who experiences them. I have tried to show that Korsgaard has a convincing

argument that the reasons of others can and must enter into our deliberative process if we are to be unified agents, if, in other words, we are to avoid having our actions appear as evil, as Hegel puts it.

The above point is a point about the nature of reason itself, the fact that reasons are available to all in much the same way. My reasons can enter into your process of reflection just the same way your reasons do. But this does not tell us to what extent justice or intersubjective deliberation actually obtains. Interpreted as a political or social problem, the question of moral luck raised by Williams, however, asks us precisely this: why is it that, given that reasons can be shared, they are *not* shared.

In concluding this chapter, I would like to take up the question of moral luck again and put it together with the idea that the constitutive standards of action change. The point I would like to make is that while Korsgaard's point is that reason is in principle public and universalizable, it is also the case that what counts as justice changes with time. Williams' notion of moral luck points to the fact that we need a historical and political account of values in order to make the possibility of public or intersubjective reason a reality.

The problem that Williams (and Hegel) puts his finger on is that injustice pervades so-called moral society because moral society is unable to accommodate individual desires. Moral society is then a set of standards or practices which is too narrow to meet the needs of self-expression of all. If we put the Gauguin example in historical context, we can see that Gauguin's problem is

both a historical achievement and the reason that norms are not yet satisfactory. On the one hand, it is already something of a success for the individualism Williams endorses that Gauguin is even able to formulate his ground project in the way he does. On the other hand, had the man Gauguin been born 500 years earlier, it might not have been possible for him to even think of leaving his family to pursue a career as a painter. So the fact that Gauguin has the option is something good on Williams' account. On the other hand, of course, the fact that Gauguin has to make the choice to leave his family in order to paint is a bad thing. It shows that there is something incompatible in fulfilling one's ground project and living with one's family, surely another ground project.

The double nature of the fact of moral luck, the contingently determined position we have in our society, suggests that something ought to be done to minimize the power of moral luck to determine our happiness. A historical account of the development of reason and reason-giving, that is, what reasons are available for us to act on, is needed in order to engage more fully with this question. I will argue in the next chapter that Hegel delivers such an account.

CHAPTER VI

HEGEL AND REFLECTION

I concluded the last chapter by arguing that Korsgaard's theory of intersubjectivity was insufficient to deal with Williams' critique of morality because, at bottom, Korsgaard can only show that intersubjectivity is possible but not that it is necessary. This, I argued, was because Korsgaard's conception of personal integrity was insufficiently developed. It is just not clear what integrity of the self is supposed to mean other than the elimination of conflict. I do, however, think that Korsgaard's notion of the constitutive standards of action provides a way of getting deeper into the problem of what constitutes integrity, not only on the personal level but on the social or intersubjective level as well.

I thus argued that the constitutive standards of action provide a theory of the intersubjectivity of reason. The fact that actions have standards which are in some sense not determined by the actor means that they are socially or intersubjectively constituted. This means that desires can be realized or reflectively endorsed in certain, socially constituted, ways. The desire to build a house is realized differently today than it was realized 200 years ago. Less obviously, though more importantly, we might also say that the way a family is structured has also changed. And we can (almost) only structure our family in ways that are currently available to us. This means that reasons for action are

intersubjectively constituted and in turn structure the agents who act on them as intersubjectively committed. “There’s nothing you can do that can’t be done”, as the famous Beatles’ line has it.¹⁷⁸

But the idea of intersubjectivity is not the only operative notion in the theory of reason giving developed by Korsgaard. As we saw, the point is also to give us some idea of how to improve our ability both to develop a more adequate set of ways of realizing our desires and a more adequate way of adjudicating between the different means to pursue our happiness both as individuals and as members of society. Such decision making depends on the set of issues which we must decide about. Korsgaard’s response to Williams is thus that he insists on the unchanging nature of individual commitments and thus allows no room for the modulation of preference or desire either historically/politically or personally.¹⁷⁹

A central thesis of the dissertation is that any account of ethical decision making must include a story about historical development and a story about how to evaluate this historical development in terms of individual reflection. Justice is the question of how to overcome the sort of impasse faced by Williams’ strong individualism and by Hegel’s man of conscience. The question of justice thus demands that the idea of reflective endorsement in the individual be expanded to include a social theory. Understanding the question of justice is thus a matter of

¹⁷⁸ The Beatles. “All you need is love”. *Magical Mystery Tour*. 1966.

understanding how we can have truly intersubjective reflection and this, in turn, is a question of seeing how the reasons on which we act are socially mediated.

In this chapter I would like to pursue the problem of the individual's place within this intersubjective realm. That means seeing to what extent the perspective of the individual itself can be accounted for within a theory of intersubjectivity. I will argue that conceiving of the process of reflection in the terms outlined in Korsgaard is the right beginning of such a theory but that it must be revised to include a theory of history. History, I will argue, is the perspective which seeks to account for those who have been marginalized by the deliberative social process and hence who form the impetus for a better theory and practice. This is the question of justice.

After this, I will turn to Robert Pippin's interpretation of Hegel's theoretical philosophy, which I believe covers much the same ground as Korsgaard's theory of intersubjectivity. This will, I hope, both expand on and corroborate Korsgaard's interpretation of Kant's theory of reflection. I will argue, however, that Pippin's reading of Hegel shows that the question of intersubjectivity necessarily leads us to a theory of history. This is so because, in Hegel, intersubjectivity is conceived of as an achievement rather than a mere possibility. Pippin's interpretation of Hegel is, however, limited by too much emphasis on the transcendental

¹⁷⁹ It should be noted, again, that Williams is much more open to the possibility of historical or political change than he is to personal change. Though, if personal change is so difficult, it is also doubtful that historical change can occur.

character of the development of thought and thus loses touch with the perspective of the subject who does not fit into the dominant conceptual scheme.

(In the next chapter I will give an alternate reading of Hegel based on the master-slave dialectic in order to show how deeply Hegel is concerned with the perspective of the injustice of dominant social norms. This reading will introduce Terry Pinkard's categorial theory, which puts the subject back into the struggle of history and justice.)

The Problem of Justice

The question this chapter seeks to answer is that of how a more detailed or determinate account of the process of reason-giving can be given. The ultimate aim of such an account is to provide a way of answering the pressing question of injustice, a question that I have been pursuing in the form of Williams' objection to the morality system embodied both by the Kantian position and by utilitarianism. In order to see how far Hegel can get us with such an account, I would like to say a little about what we might be looking for in such an account. The beginning of an answer is given by Pippin in a paper criticizing Robert Brandom's appropriation of Hegel for being insufficiently attentive to the difference between subjugation and fully developed, that is, just, social interaction. Brandom's holism, as it is portrayed by Pippin, bears some resemblance to Korsgaard's theory of intersubjectivity in that it too develops a type of social autonomy of communal deliberation.

Pippin summarizes Brandom's inferential holism as claiming that "what commitments you undertake are up to you but the content of those commitments, just what you are committing yourself to by committing yourself to claim P, is not; that is 'administered' by others. ('I commit myself, but then they hold me to it')." ¹⁸⁰ In Korsgaard's terms, we could render this argument by saying that if I allow you to enter into my process of reflection, I must take your reasons as normative constraints on me. This account, Pippin argues, is not so much wrong as incomplete, since it does not seem to distinguish between the normativity of the real historical institutions we are being held to and the normative development that comes from contesting those historical institutions. For Pippin, this means that Hegel's project is tied up with the claim of being able to make sense of the actual continuities and crises of history during any attempt to create (just) institutions. This means that Hegel wants his account to say something about the justice of any challenges to power. Brandom's view cannot do this since it cannot say anything beyond that there appears to be controversy in people's reasoning practices. Brandom remains a historical sociologist as opposed to a critical theorist of morality. ¹⁸¹

Hegel, Pippin contends, agrees with Brandom in the sense that meaning is a matter of use, but disagrees with Brandom about what we can therefore say about meaning. For Pippin, Hegel wants to be able to say why certain articulations of norms failed, something Brandom does not appear to be

¹⁸⁰ Pippin. "Brandom's Hegel". *European Journal of Philosophy*. 13:3 (2005) : 381-408. 391.

¹⁸¹ Pippin. "Brandom's Hegel". 392.

interested in. I think there is a similar lack of interest in the historical structure of the constitutive standards of actions in Korsgaard's work. For Pippin's Hegel, the way the Greek notion of citizenship failed tells us something about what citizenship and authority actually is.

This distinction between the historical normative approach and the inferentialist approach comes down to being able to see history itself as a critical and normative process. History, for Pippin's Hegel, is thus both a slaughter bench and a source of meaning. Brandom's account of interacting individuals, Pippin writes, "will not even allow the problem that bothered Hegel his entire career to arise: that problem of 'positivity', subjection by others, according to appropriate, public practices, to a status of 'undertaken commitments' not recognized as such by the individual."¹⁸²

Pippin concludes:

The whole ball game in Hegel comes down to the question of whether he has in fact discovered a historical, developmental way of making the case that this distinction [between subjugation and a form of life in which the freedom of one depends on the freedom of all] can be made (without any form of moral realism or Kantian 'moral law' universalism), of saying what institutional form of life actually achieves these desiderata, and his being able to show that it is the unfinished and still unfolding achievement of modernity to have begun to do all this.¹⁸³

I will argue below that on Pippin's own account, Hegel does not succeed in this project because he is unable to give an account of the rationality of historical categories. And absent such an account, as Pippin readily

¹⁸² Pippin. "Brandom's Hegel". 395.

¹⁸³ Pippin. "Brandom's Hegel". 397.

acknowledges, Brandom's inferential holism has much to recommend it.¹⁸⁴ But the more general point to be taken away from Pippin's critique of Brandom, however, is that a theory of normativity must be able to say something about what it is like to be an agent within the normative system. It must be able to articulate the subjectivity of the agent involved in deliberating with others. Absent such an account, we will not be able to say much about how the system is to be improved. This is the point pressed by Williams: how is it that we must continue our allegiance to a morality system which continues to subjugate the desires of individuals to the desires of 'all' (where 'all' remains vague)? The question of subjectivity is thus also a political question.

But the question of subjugation raised by Pippin also reveals something about the historical process linked to the notion of justice. Subjugation, passivity in the face of the power of others, is only the flip side of agency and freedom. So by focusing on the question of what it looks like for the agent to be within the social system, Pippin also reminds us that the social system is not some monolithic mechanism of repression (as Williams sometimes portrays it) but the interplay of agency and passivity. History itself is then the story of how the disenfranchised try to arrange things in a better way. The question of justice is then whether the agency of each of the individuals involved does something to further the satisfaction of each of the other individuals involved in the society in a

¹⁸⁴ And, we might add, to take up the question of overcoming metaphysics once again, absent the speculative grounding of normativity, Brandom's view of 'score keeping' does not really differ much from Rawls' coherentism.

non-arbitrary way. That is, to what extent does the satisfaction of one person's needs involve others.

The possibility of what we might call collective agency, undeveloped as it is, reminds us that society is made up of the interaction of individuals who have spontaneous desires rather than agents who act out pre-given or pre-assigned roles. The role of a historical normative theory must thus seek to make sense of the ways in which people continue to break out of pre-assigned roles. Agency, as both Kant and Hegel insist, must be conceived of as radically free. This is because it originates not in the social world against which our struggles sometimes appear as futile, but in consciousness itself. The freedom given by our capacity to reflect is the origin of the historical determination of norms and also its limit. In seeking a theory of justice, we would do well to avoid both horns of the dilemma: *underdescribing* what it is like to be an agent *and overdescribing* the set of norms developed by consciousness by making the norms too determinate. For in order to be binding on us, normativity must be both freely endorsed and also binding. But the individual is the only one who can determine the norm's force. The question for Pippin's interpretation of Hegel thus becomes: how are concrete social relations possible and rationally justifiable? In order to see whether concrete social relations can be rationally evaluated, we must examine what, in Pippin's interpretations at least, Hegel takes to be the rational conditions of thought. This will take us further into Hegel's speculative philosophy.

Hegel's Appropriation of the Transcendental Unity of Apperception

In what follows, I present Pippin's interpretation of Hegel's theory of reflection. I believe this will help us see several points: the first point is that Korsgaard's theory of reflective endorsement is in many fundamental respects quite similar to Hegel's, since both originate in Kant's theory of apperception. Pippin's account of Hegel can thus be seen as an expansion on Korsgaard's view of self-consciousness, a part of her theory which she does not spend much time on. The second point I would like to make is that there are general argumentative similarities between Korsgaard, Rawls and Hegel. All three, I think, argue that there is only one principle of action, the categorical imperative (for Korsgaard and Rawls) and the idea of the self-development and recognition (in Hegel). Presenting the Hegelian theory here will allow us to see that by arguing that the hypothetical imperative is fundamentally subordinated to the categorical imperative, Korsgaard shows that all actions are normative, and more importantly that there can be no stable distinction between the 'moral' and the 'non-moral'. Pippin's discussion of the development of self-consciousness in Hegel will show that for Hegel too, what might be called instrumental reasoning with regard to objects (or epistemology in general) is really subordinate to the development and interaction between agents. If there is something like a categorical imperative in Hegel, it is the imperative to develop an ever more adequate way of interacting between agents.

Lastly, I will link Hegel's imperative to justice and will end this chapter with a criticism of Pippin's interpretation of Hegel's master-slave dialectic. Pippin, I will

argue, does not conceive sociality in sufficiently dialectical terms and so reifies the notion of justice which he formulated in discussion with Brandom.

Reflection

Let us now turn to Hegel's theory of reflection and its origins in Kant. For Pippin, the origin of Hegel's speculative philosophy, and thus the key to understanding how Hegel sought to overcome what he perceived as Kant's metaphysical baggage, lies in Kant's deep insight into the role that consciousness plays in developing our conceptual schema. Pippin argues that Hegel's indebtedness to Kant is far greater than Hegel ever admitted and might be put as follows:

Keep the doctrine of pure concepts and the account of apperception that helps justify the necessary presupposition of pure concepts, keep the critical problem of a proof for the objectivity of these concepts, the question that began critical philosophy, but abandon the doctrine of "pure sensible intuition," and the very possibility of a clear distinction between concept and intuition, and what is left is much of Hegel's enterprise.¹⁸⁵

For Kant, in making a judgment I must be able to recognize myself as the one making the judgment, for otherwise it would not be me making the judgment. As Pippin puts it, being *in* a subjective state, even a merely momentary subjective state, does not count as having an experience *of* and so being *aware*

¹⁸⁵ Pippin. *Hegel's Idealism*. 9.

of that state unless I apply a certain determinate concept and judge that I am in such a state, something I must *do* and be able to know that I am doing.¹⁸⁶

Awareness of my judgment is not something I happen to have but the condition of the judgment being mine. This means that the judgment is reflective.¹⁸⁷

This reflective judgment means for Kant that the 'I' of apperception must supply a form for the judgment, and since only certain forms are available to us, only certain types of judgments can be made. These are the categories of the understanding. But this Kantian claim raised the following problem: How should one understand the claim that 'intuitions must conform to the categories for experience to be possible'? Does this claim amount to a demonstration of the objective reality of pure concepts (as Kant thought) or does such a claim amount to the assertion that we can know a priori that intuitions conform to categories because there is no real independent 'givenness' in experience, and thus that an 'identity' between concepts and intuitions has been established, that 'thought' has successfully determined its 'other'?¹⁸⁸

The use of Kant's apperception thesis is supposed to avoid the realist challenge according to which, even if *our* best criteria for 'knowledge of X' are fulfilled, we still have no way of knowing whether such fulfillment does tell us

¹⁸⁶ Pippin. *Hegel's Idealism*. 19.

¹⁸⁷ See also Pippin's working out of this problem with regard to Kant's *Critique of Judgment*. "Avoiding German Idealism: Kant, Hegel, and the Reflective Judgment Problem". In *Idealism as Modernism*. pp 129-156.

¹⁸⁸ Pippin. *Hegel's Idealism*. 30.

anything about X. Realism is supposed to be undercut by claiming that the only possible knowledge we can have stems from our subjectivity.¹⁸⁹

Without examining the details of this view, we can see that Hegel's reinterpretation of the Kantian apperception thesis sets the stage for a more ambitious theory of 'what there is'. By rejecting the notion that there is a distinction between receptivity and spontaneity, and therefore rejecting the idea that there is something beyond the reach of reason, Hegel is poised to construct a theory of reason which develops itself purely out of itself and must therefore justify itself purely on its own terms.

According to Pippin, Hegel is thus committed to a theory about the priority of pure concepts in human experience. Hegel argues that the Notion (*Begriff*) determines the possibility and character of human experience and that, since there is no contrast between our conceptual framework and the world itself, the Notion does not limit us.

Pippin's argument about Hegel thus also suggests that Hegel is trying to overcome the skepticism which was implicit in Kant's (and Fichte's) position.¹⁹⁰ By moving now to the *Phenomenology*, we can see that the issue that must be overcome will be the skeptical challenge according to which there might be a gap between self-certainty and truth. In answering the skeptic, Hegel turns the table in a radical manner by arguing that skepticism arises in a determinate or concrete way only as a result of the truth of the Notion itself. Pippin writes:

¹⁸⁹ Pippin. *Hegel's Idealism*. 39.

¹⁹⁰ Pippin. *Hegel's Idealism*. 91.

[Hegel] will try to show, determinately, why, given some putative Notional determination of objects, doubts about whether objects must or even can be so Notionally specified, are the relevant, determinate doubts they are, only as a consequence of that Notion's own incompleteness. This in turn means, for Hegel (summarizing everything at once), that such an "opposition" between "subject" and "object" is itself a "determination of the Notion," and so such an incompleteness can itself be made out only on the assumption of a developing Notion of objectivity.¹⁹¹

According to this programmatic claim, the first three chapters of the *Phenomenology* can be seen as a step by step refutation of different sorts of skepticism about the empirical world, concluding that consciousness is indeed capable of understanding the world as it is. However, even if the apperception thesis can refute skepticism about knowledge of the external world, this still does not tell us how the world actually *is* organized. Much will depend on whether Hegel can show us how we do understand the world. In rough terms, Pippin thinks that Hegel's development of notional adequacy vis-à-vis the world is underwritten by a dialectical theory according to which state B is the resolution of state A's difficulties until a better candidate comes along.¹⁹² Much will depend on how the apperception thesis is turned into a thesis about conceptual determination and hence into a theory about how the structures of the world make possible our interactions with each other. This is the question of recognition. The point which I would like to insist on is that the critique of metaphysics presented in the first three chapters of the *Phenomenology* is in the

¹⁹¹ Pippin. *Hegel's Idealism*. 98.

¹⁹² Pippin. *Hegel's Idealism*. 108.

service of a theory about how people can interact with each other and what mediates peoples' interaction with each other.

Self-consciousness and the Apperception Thesis

For Pippin, Hegel's path toward a theory of recognition or concrete social determination begins in the fourth chapter of the *Phenomenology*. The chapter starts with the idea that, given the apperception thesis, the subject must be conscious of the fact that it is differentiating objects when it does so.

[Kant and Hegel both assume] that the results of any natural or causal relation between an object and a sensory episode in me could not count as my representing the object unless I take myself to be representing that object, unless the object is "for me" the object of my representing activity, and this self-conscious activity requires an account of its conditions.¹⁹³

The apperception thesis, however, at least in its Kantian version, suggests that there is a difference between the spontaneous work of the understanding and the receptivity of the world outside the subject. And it is this thesis that Hegel must now make his own by arguing that the subject does not meet objectivity when it gazes outside of itself, but rather its own subjectivity. That is, instead of the Kantian account of the limits of knowledge, Hegel develops an account of how, as Pippin puts it, "we ought to be satisfied that the ways in which the world

¹⁹³ Pippin. *Hegel's Idealism*. 132.

and others are taken up and assessed are well grounded or ‘absolute’”.¹⁹⁴ In *Hegel’s Idealism*, Pippin characterizes this movement as leaving behind the deduction of the categories of the understanding in the first three chapters of the *Phenomenology* and embarking on a full blown anti-realist pragmatism in which “Hegel appears to be saying that the problem of objectivity, of what we are willing to count as an objective claim in the first place, is the problem of the satisfaction of desire, that the ‘truth’ is wholly relativized to pragmatic ends.”¹⁹⁵ Hegel is thus claiming that knowledge is a function of human interest. What is at stake in the rest of the *Phenomenology*, then, is a historical development of some form of ‘practice’ which gives us institutions which reflect our like-mindedness. Human interest, in other words, is social interest. This claim involves considering what Hegel means by ‘mutual recognition’ and how it is related to Hegel’s claims about the adequacy of human conceptuality or absolute knowledge.

The move from the deduction of the forms of our understanding to the development of the forms of human ‘life’ in general to the forms, in other words, of human reason, requires a transition from the paradigm of self related to world, to self related to self or, as Hegel has it, from consciousness to self-consciousness. The move from consciousness to self-consciousness is also the move from the ‘I’ to the ‘we’ and is thus the beginning of intersubjectivity.¹⁹⁶ Since

¹⁹⁴ Pippin. “You Can’t Get There from Here: Transition Problems in Hegel’s *Phenomenology of Spirit*”. 61.

¹⁹⁵ Pippin. *Hegel’s Idealism*. 148.

¹⁹⁶ Though Pippin does considerable work to show that this transition is not quite as stark as it has appeared to some commentators, it should also be noted that this move is really enormous in

Hegel has already repeatedly denied that objects affect us immediately in the form of some sort of 'given' and claimed that our interaction with the world is structured by the concepts we employ to make sense of it, it now appears that in attending to the world we are actually attending to our own mode of perceiving the world in perceiving it. That is to say, in comporting ourselves in the world, our comportment becomes an issue for us (to sound a little Heideggerian).

Given this 'epistemological' sounding claim, we must do some work in order to account for the seemingly sudden introduction of the language of desire at the beginning of chapter four of the *Phenomenology*. Pippin makes the transition using the concepts of dependence and independence. What might help us here is to see that in developing an account of human interaction, Hegel is aided by a point already implicit in the apperception thesis: that normativity is pervasive rather than something attained at a particular level of social interaction, such as morality. Self-consciousness presents us with the paradigm of autonomy and hence of normativity. In reflective endorsement we commit ourselves to beliefs and to actions in the same way. Independence and dependence on other people are thus the paradigm concepts of autonomy itself.

A living subject might be relatively independent of the natural world by being in a 'negative' relation to it, that is, by overcoming obstacles and consuming parts of the world, but this means it is also dependent on the world

terms of the history of philosophy. Hegel here appears to undercut the idea that there is an ontological distinction between apperception of a subject and apperception of an object. The

because something other than the subject's own 'self-consciousness' is its essence. In order to understand what it means to be 'self-conscious', we will have to render determinate our desires and see to what extent they are indeed related to our relation to ourselves.

Hegel's claim here is that relations to objects do not satisfy us because they do not establish any sort of permanent relation with us. Something like self-constraint or self-relation only occurs when the subject takes objects *to be* objects of its desire and thus worth pursuing. But for this to be the case, the object must be permanent. And, Hegel adds, a desire is only a desire for the subject in a self-conscious way when there are two desiring agents whose desires clash.¹⁹⁷ This means that it is only possible to conceptualize the self as a self via another self—only the other self can give sufficient resistance to become the object of a steady relation.

We might also understand the other's resistance to the subject as providing the first subject with a more determinate category for understanding the world. In this sense, the appearance of the other is actually the appearance of a new category of understanding. Such categorial self-understanding has a historical horizon which constitutes sociality itself. It is no accident that the moment of history appears along side the moment of subjugation but also alongside possible equality or recognition. For it is the struggle for the promised

speculative point in the apperception thesis is thus to show that inanimate objects and human beings exist on a continuum of categorial development.

¹⁹⁷ Pippin. *Hegel's Idealism*. 151-52.

equality, which the subject recognizes but then rejects in the other, that is the stuff of history. The point is that dependence does not have to mean subjugation but can also be understood as equality, as interdependence.

The Master-Slave Dialectic

Pippin contends that rather than providing a social ontology or anthropology in the famed master-slave dialectic, Hegel is proposing a way of understanding the relation of self to self in terms of an elaboration of the apperception thesis. In the social realm, the apperception thesis turns into the autonomy thesis discussed in previous chapters.¹⁹⁸ The problem with the independence of the subject from the natural world is that it was indeterminate. The world yielded too readily to the appetites of the subject and so could not furnish any sort of self-understanding. This is the same point made earlier about the private language practitioner. No one was there to hold her to the rules that she claimed to adhere to. The problem is that of solipsism. To fight solipsism we need an object that can and wants to assert its independence against the subject. Hegel conceives of this encounter between subjects as a state of war, because they share no common ground. Each desires merely its own self-relation, that is, its independence. Independence gains its value from the specter

¹⁹⁸ Since Hegel's argument does not depend on any sort of proof or deduction, but rather on the speculative and historical thesis that reason develops out of itself, we cannot make any distinction here between the 'weak' and the 'strong' autonomy thesis. Indeed, Hegel's developmental strategy must be seen as a response to the inability of a Kantian deduction of reason.

of dependence which has just arisen; there is now someone who might be able to hold the agent to her own standards. But since both desire the same thing, desire is mediated and turned into knowledge. Pippin writes:

Knowledge of objects is conditioned by forms of self-consciousness, and forms of such self-consciousness are to be understood as the product of opposed self-consciousnesses attempting to resolve such opposition, ultimately in 'thought'. 'Recognition' is Hegel's name for the achievement of such collective subjectivity.¹⁹⁹

There is, of course, a paradox in this struggle for collective subjectivity. Though it is ultimately supposed to issue in social recognition, the struggle initially ends in the radical subjugation of the slave by the master. Knowledge as social justice has not been achieved. Hegel finds the true self-understanding, resulting from the struggle to the death, in the notion of work. For in working, the slave recognizes that there is no independence to human activity and that human activity is always bound by contingency and death. But this, in turn, means that any sense that *can* be made of one's life arises wholly as a result of and within the self-defined experience of the collective activity of labor. It is by laboring, transforming the world in the service of the Master— a labor that requires a determinate understanding of nature, and of knowledge itself— that truly independent and so self-determining self-consciousness arises.²⁰⁰ Dialectically, this means that the master's independence from the material world is not true freedom at all. True freedom, Hegel argues, exists only in the adequate mediation of self-positing freedom, which includes accepting social parameters.

¹⁹⁹ Pippin. *Hegel's Idealism*. 160.

True freedom, in other words, is when we have concrete categories through which to interpret the world and ourselves.

The prospect of determinate categories of concrete social relations raised here by Pippin is, as we shall see, controversial. This interpretation of Hegel goes back to Pippin's earlier claim that the answer to the struggle between the two individuals resulting in the master and slave positions is ultimately provided by thought itself, and more specifically in Hegel's *Logic*. Let me put the worry about this point thus: If the struggle for recognition can indeed be carried out in thought, then it appears that only thought will determine the adequacy of each successive step toward recognition. But, as we have repeatedly noted, the perspective of the combatants, be they Gauvain and his family or the master and the slave, is also determined by contingency, the structures of desire that the individual just happens to have. So, the potential problem with developing a theory of recognition out of thought alone is that it will not be able to take into consideration the perspective of the agents involved in history itself and will thus only be able to form an abstract notion of justice rather than the more concrete one called for by injustice.

Self-Positing Thought and Contingency

H. S. Harris, commenting on this chapter in Pippin, has added an interesting interpretation. Harris argues that Pippin has shown the necessary

²⁰⁰ Pippin. *Hegel's Idealism*. 162.

dual nature of freedom by contrasting the life of the master and the life of the slave as he does. Chapter four of the *Phenomenology* is so central to Hegel's general argument because it attempts to resolve a tension in the argument that has existed within the Kantian version of the apperception thesis, namely that there is a decisive difference between self-consciousness and self-reflection. While the former is the awareness of oneself as judging, the latter is what issues from the judgment itself, a judgment about objects.

The master-slave dialectic illustrates both the difference between these two and the fact that both are necessary in order to understand Hegel's notion of freedom. On Harris' analysis, the problem with desire is that I don't even know what I desire until I have desired it in struggling for it. This means that Hegel (and Pippin) are right to conceptualize the attainment of self-consciousness in the struggle to the death. Self-consciousness is blind desire, it wants only its own self-assertion, it thus posits pure freedom. Reflective freedom, coming afterwards, chooses a relation among objects. This further fleshes out the requirement that the apperception thesis yield concrete categories through which to live our lives.

Harris' point is that we must understand both types of freedom if we are to understand what Hegel means by self-determination:

'Self-consciousness' is a *spiritual* relation. At the minimum (in that very boundary situation that reveals its logical primacy) self-conscious freedom involves a double consciousness: the one that is 'self-consciously' aware of freedom, and the one that is 'consciously' (or 'reflectively') aware of it. *Both* consciousnesses are 'free'— but one of them has *asserted freedom*, while the other has *freely accepted* the 'lordship' of death. Kant and Fichte have one half of the truth about the self-positing of Reason, Hobbes has the

other. *Neither side* can successfully explicate the *actual* rationality of experience. The choice— to live or to die— is radically free . *Neither* choice (by itself) is 'rational'. The existence of 'Reason' requires that both choices should be made together. A *singular* self *cannot* 'posit itself freely'.²⁰¹

The experience of choice and the paradigmatic result of the master-slave dialectic might thus be formulated as “In choosing freely I must accept as mine what results from my choice.” Harris’ point shows that even in formulating his anti-realist pragmatics, Hegel has not given up on the idea of contingency or a certain type of ‘receptivity’ and ‘spontaneity’. The active and passive have simply moved to the social level. Harris fleshes out this thought out by giving the example of the soldier going off to war. While the soldier’s mother is quite reasonably expected to hope that her son doesn’t die, *he* is not supposed to wish that. Or rather, though it is reasonable for the son not to wish to die, it is his willingness to die which makes *him* truly rational. The social dimension of this division can be seen in the fact that different reactions to the event of going to war are now distributed among different people. The only thing that can make sense of wanting to go to war and being prepared to die is that part of the soldier’s self resides outside of him, in his mother, say.

Pippin’s own gloss on the importance of the separation of the apperception thesis into Kantian self-consciousness and more mediated self-reflection is to argue that Kant left the subject ‘homeless’, undirected and alienated.²⁰² The task of the *Phenomenology* and indeed of the rest of Hegel’s

²⁰¹ H. S. Harris. “The Problem of Kant”. 23.

²⁰² Pippin. *Hegel’s Idealism*. 165.

system is to understand how the subject can retain its pure freedom within the contingency of the material world, that is, remain radically free in thought even though his body is shackled.

The idea that there could be two moments involved in the development of freedom might, however, cast some doubt on the project of developing a determinate scheme for conceptual development, since it presupposes that part of the work of freedom is accepting contingency. If the contingent is really radically contingent, as it must be, then we are always dealing with different ways in which history must be made free. A theory of freedom which insists only on the first sort of freedom, self-positing, is in danger of missing this important point.

At this point, Pippin's argument in *Hegel's Idealism* turns to the *Logic*, and we cannot follow it here. Let us note, however, that by turning to the *Logic*, Pippin implies that the *Logic* works out the conceptual schema of the individual self-consciousness. This implication seems to be at odds with Pippin's interpretation of Hegel's claim that the *Logic* develops a determinate structure of thought uncoupled from individual subjectivity.

Linking Hegel's Speculative and Practical Philosophy

Given the objection to the idea of developing a social theory through thought alone, the question for Pippin's interpretation of Hegel is whether he can indeed link his compelling speculative thesis about apperception to a robust thesis about autonomy and justice. The success of this task will depend on whether Pippin's Hegel can give a strong account of the necessity of those social

institutions which are ultimately supposed to mediate our intersubjective relations. Such a theory will not only have to argue for the necessity of those institutions but will also have to show how these institutions provide the normative foundation for our actual social life. Pippin will thus have to provide us with a set of criteria enabling us to accept some particular instantiations of the state and reject others. The general question, then, is whether Pippin's Hegel can translate the apperception thesis into a theory of autonomy or freedom which can both help us be satisfied with the life we do live and also demand change where our current form of life is inadequate. Kant, in Hegel's critique at least, could not help us with this problem because the principle of pure practical reason Kant developed was inadequate to mediate between subject and world. We thus have to look for a very deep notion of reason and justification in Hegel.

Pippin claims that the proof is in the pudding, that is, that Hegel's social theory must justify itself by providing a lucid analysis of modern structures of authority. Pippin himself provides no philosophical link between the speculative part of Hegel's philosophy and the analysis of social structures.²⁰³

Pippin interprets Hegel's social theory using the metaphor of a self-grounding. I call this a metaphor because Pippin never explains this grounding in his writings on Hegel's practical philosophy. This suggests to me that Pippin actually finds very little in Hegel's practical philosophy which displays sufficient

²⁰³ Pippin even doubts that such a connection can be established. See *Hegel's Idealism*. 258-60.

internal rationality to ground itself.²⁰⁴ And this, in turn, suggests that Pippin sees serious problems for Hegel's practical philosophy if it is to provide more than just another model of sociality, like *A Theory of Justice*. The language of self-becoming does help Pippin distinguish Hegel's view from a number of other views. Hegel's practical philosophy, therefore, is not a form of communicative action, as Axel Honneth contends.²⁰⁵

The repeated insistence on 'becoming' or 'self-actualization' has led Richard Rorty to complain that Pippin's account serves only to celebrate autonomy rather than to explain it. Rorty claims that to say humans have become free in modern society, absent some sort of causal explanation, is like saying of an author 'this is the book you were born to write'. Such Aristotelian potency-act distinctions, Rorty writes, now merely strike us as celebratory and at worst as

²⁰⁴ Pippin gestures to the family as a possible way of developing a self-grounded social structure. "In the simplest terms, everyone has parents, can reproduce the conditions of their existence only cooperatively, and are invariably subject to, or the subjects of, decisions about the common good or the exercise of some sort of political power. We are not simply one agent among many, or all alike in being agents who can act on reasons. We are, but even in being able to recognize and act on such considerations, we require others, such that the socially formative and educational institutions which make possible such recognition and its realization are effective." Pippin. "Hegel on the Rationality of Ethical Life". 123.

²⁰⁵ This is so because communicative action already assumes the existence of rights which are not fully realized, however, until the rational state is. "Hegel and Institutional Rationality". 10-11. Nor is it Aristotelian or Kantian. The account is not Aristotelian because it does not presuppose any substantive notion of the good, though Hegel's account is, in some sense teleological. "Hegel on the Rationality and Priority of Ethical Life". 121-22. Hegel's is not a Kantian account in the sense that it breaks with Kant's individuated ontology of the subject.

merely verbal.²⁰⁶ Given the seriousness and philosophical rigor with which Pippin has pursued Hegel's thought, Rorty's characterization seems rather reductionist of Pippin's project, but it does strike at something. I would like to propose that we understand Pippin's reluctance to concretely articulate the rational principles behind Hegel's notion of ethical life as having to do with Pippin's final judgment on Hegel's speculative project itself.

The Transcendental Thesis and Hegel's Failure

Pippin cannot flesh out what the theory of institutions concretely means because of what Terry Pinkard has called Pippin's transcendental thesis: the idea, taken from Kant, that we need to construct categories without which we could have no conceptual schema at all.²⁰⁷ According to Pippin, these categories must be the product of determinate negation which means that they have to be of the same sort as the categories of ordinary logic. Dialectical logic, however, cannot accomplish this task, as Pippin points out. Indeed, the problem of Hegel's dialectic is that "The 'urge' that Hegel speaks of for the Notion to comprehend itself in these [dialectical] moments is [...] an urge Hegel cannot demonstrate even if he presupposes it throughout."²⁰⁸

²⁰⁶ Richard Rorty. "Comments on Robert Pippin's 'Naturalness and Mindedness: Hegel's Compatibilism'". 216.

²⁰⁷ Pinkard. "The Categorical Satisfaction of Self-Reflexive Reason". 6.

²⁰⁸ Pippin. *Hegel's Idealism*. 255.

This lack of dialectical rigor means that the categories of reason which Hegel does develop are not truly self-grounding at all. Taken together with Pippin's own reservation about the possibility of translating logical analysis into the language of social philosophy, this means that no satisfactory account of the self-grounding nature of the categories of social life can be given, and that Pippin must remain vague about their rationality. There is no way for Pippin to give an account of what it truly means for freedom to be expressed in the thesis that "In choosing freely I must accept as mine what results from my choice." There is no way to explain the unity of the apperception thesis (the self-positing of freedom) with its reflection in social and historical categories according to the transcendental thesis.

Hegel's failure to live up to the transcendental thesis thus means that Hegel fails to give a rational account of ethical life, whatever else he might do that is helpful in our understanding of sociality. These latter insights Pippin has charted with great clarity. For Pippin, however, Hegel is ultimately unable to give a speculative grounding to his social philosophy in much the same way that Kant is unable to: Hegel is unsuccessful in showing that the categories the subject develops are in fact the categories which reflect the totality of actual relations in the world. And, lacking the certainty that our categories are right, we do fall back into some sort of skepticism.

Certainly, Pippin has formulated Hegel's project in the most demanding terms. Hegel is held to his avowed intention to furnish us with a complete theory of what there is. The requirement of determinate negation in Pippin's Hegel

suggests that the theory of what there is be given for all time and thus be exhaustive. But the ambitiousness of Hegel's project seems ultimately to betray the very thing Hegel (and Pippin) have their eyes set on, namely, a theory of justice. And it is to Pippin's credit that he does not try to spin a theory of justice out of a metaphysical theory of the conceptual schema. To do this would not only be to betray the ideal of freedom, but actually to violate it.²⁰⁹

Conclusion

Lastly, let me formulate the problem with Pippin's interpretation of Hegel in a way that will allow us to see its link to Korsgaard as well. In Pippin's interpretation, Hegel is fundamentally interested in showing the objective structures of conceptuality. This means giving an account of sociality as it exists in its current and complete form. The problem is not only that such an account, as Pippin recognizes, is not possible, but, as Pippin does not recognize, such an account would be too complete to be true.²¹⁰ By this I mean that if a complete account of intersubjectivity could be given, this would mean the exclusion of the perspective of those whose views are in conflict with the dominant account as well as the perspective of the speculative theorist herself.

To take up the theme of the master-slave perspective, we could say that on both Korsgaard's too general view about the possibility of intersubjectivity and on Pippin's too determinate view, the perspective of the slave cannot come into

²⁰⁹ I take it that this is the criticism Adorno levels at Hegel's project. *Drei Studien zu Hegel*.

view. This means that neither view contains the necessary relation to the contingent world which would force it to develop an account of the concretely unjust relations which exist within the potentially just social world.

What I have thus been trying to show in this chapter is that the project of developing a way of understanding the social world through the process of reflection itself is doomed to fail if it does not also include a dialectical relation to that which it is, at all times, trying to assess: the material. The problem with Pippin's view of Hegel is thus what is also fundamentally the problem with Kant's view, namely that neither Kant nor Pippin's Hegel seem to believe that nature or contingency affects our freedom. In seeking to rescue our freedom from determinism, both represent nature as fundamentally cut off from our unfree historical existence.

This is why, I want to argue, Pippin's interpretation of sociality actually lacks a theory of history. History is tied to injustice just the way Pippin says, namely by being the story of how inadequate institutions are challenged by individuals and cultures who are not represented by them. But the story of these challenges, because they are themselves contingent, cannot be represented in logical thought. They require a more fluid story which takes the perspective of the subject within history without being able to also give it a transcendental structure.

²¹⁰ I mean this in the sense of Adorno's dictum: "Das Ganze ist das Unwahre." *Minima Moralia*.

CHAPTER VII

HEGEL, INTERSUBJECTIVITY AND HISTORY

In this chapter I take up the problem of justice by developing an interpretation of what one might call Hegel's moral psychology. By taking another look at the master-slave dialectic I would like to show why Hegel thinks that the problem of intersubjectivity or recognition is intimately connected to the problem of giving a historical account of justice. I will argue that Hegel conceptualizes the path toward justice as incremental improvement both in the rationality of the actual choices that individuals make and in the range of choices available to them. What drives the dialectic is the constant dissatisfaction with the current set of options each individual and the culture as a whole experiences. History, in Hegel, is thus conceptualized just as much by the moving away from injustice as the progress toward justice.

In the second half of the chapter I will turn to the work of Terry Pinkard in order to take up some of the questions left over from the previous chapter. I will pay particular attention to how Pinkard proposes to overcome the need to read Hegel as a transcendental philosopher. Pinkard's category theory approach, I will argue, preserves the immanent characteristic of Hegel's philosophy by insisting that Hegel always views the development of justice and recognition from the perspective of its incompleteness. The perspective of the incompleteness of

justice, I will argue, is the perspective of the subject trapped within a system of ethics which it can only partially condone and respect. This perspective makes it incumbent on us, as participants in history, to be open to and act to further revisions of the current ethical status quo.

The Master-Slave Dialectic: Radical Freedom and Slavery

I would now like to say a little more about the master-slave dialectic in Hegel in order to provide a way of avoiding the problem Pippin's interpretation of self-consciousness faced. The main point I would like to make here is that, in the master-slave dialectic, Hegel develops a stronger notion than Pippin's account admits of how consciousness and the material world collide and of the damage that this collision inflicts on individuals. The master-slave dialectic, though concerned with the possibility and also the need for just interaction or recognition, is also the place where Hegel begins to articulate the possibility of interaction as a historical achievement and so as the result of deliberate action by individuals. It is also where Hegel introduces a notion of agency which is quite close to Korsgaard's. The aim of this section is thus to put the notion of injustice and justice in relation to a historical narrative of individual and social agency.

Let me first say something about Hegel's notion of equality and difference and then proceed on to a discussion of how the concept of labor (*Arbeit*) unifies the subject and sets it on its historical path toward justice.

Hegel begins the master-slave section by announcing the intended result of this dialectic: "Self-consciousness exists in and for itself when, and by the fact

that, it exists for another; that is, it exists only in being acknowledged”²¹¹ By “in and for itself” I take Hegel to mean something akin to Korsgaard’s notion of having integrity, not being a divided self. Recognition is a matter of being a unified or integrated self among other unified or integrated selves. I think that Hegel would accept much of Korsgaard’s account of the necessity of thinking of agents in this way— as at least possibly unified with others and yet autonomous— so I do not think we need to go over the argument again.

What is of concern for us is rather why Hegel thinks that an account of recognition must be historical. Hegel’s answer is that we need a historical account because it is only in the concrete interactions between agents and social norms that agents and social norms are even constituted at all and hence can even begin to build a just society. In order to understand this point, we must understand why Hegel thinks that justice appears on the historical scene first as injustice.

Here I want to come back to a point made earlier by Harris. Harris noted that the process of thought in general really has two sides. The first side is the radical posited nature of freedom. Harris called this self-consciousness. The other side is reflection, which is the endorsement of one desire over another. Hegel wants to elucidate the interaction between these two moments of thought by placing the emphasis not on the positive aspect of recognition, i.e. the idea

²¹¹ Hegel. *Phenomenology*. 111. “Das Selbstbewußtsein ist an und für sich, in dem, und dadurch, daß es für ein anderes an und für sich ist; d. h. es ist nur als ein Anerkanntes.” *Phänomenologie des Geistes*. 109.

that recognition is in principle possible because the structure of each self-conscious self is the same, but on the negative, on the inequality of the two self-consciousnesses. I think that we should view these two sides of the self as being that of the radically free self and the other as that of the unfree because embodied self. Hegel, in other words, wants to see what happens to the self of self-consciousness when it runs up against contingency, and this is precisely what it does when it is thrown back onto its other.

Coming upon its other triggers a reorientation of the self toward the world— a world which has hitherto not existed because nothing in the non-human material world resisted the self on a conceptual level quite in the way the other does. Certainly, securing food and shelter provided certain ‘technical’ challenges, but these were either overcome or left behind. All activity was still in the fundamental service of the agent’s radical freedom, its self-consciousness or self-positingness. It is only with the appearance of the other (or the doubling of the self) that an obstacle appears which threatens the subject’s pure self-consciousness. The subject thus takes a stand, insisting that its self-positing self-consciousness is more important than its contingent mediation of the objects of the world by desire. The subject risks its life. This is because the mere existence of the other’s radical freedom is a threat to it.

It is important to see that the other becomes a threat to the self by being quite literally *there*, by appearing in (for the first time) social space. Social space is fundamentally a limitation of the self’s ability to pursue its desires. Social space thus threatens and enables the self’s radical freedom. Implicit in this analysis is

the worry that the self might break apart, that the self's radical freedom might be replaced by the other self's radical freedom and yet that the first self's process of reflection would continue to exist. This, I think, is what Hegel has in mind when he speaks of masters and slaves.

What makes the slave a slave is that its ability to pursue its 'own' reasons no longer exists. The slave's ability to have desires, its radical freedom, has been replaced by desires coming from the outside, from the master. The worry is, in other words, a worry about the possibility of some sort of radical externalism. The internalist position, as we saw, says that you can only act on desires which you *yourself* have. The Kantian interpretation of this was to say that you must reflectively endorse your desires in order to turn them into reasons which are normative for you. This is how you act or at least intend to act. The existential threat the other poses is not on the level of reflective endorsement, but rather at the more fundamental level of having desires in the first place. For it is radical freedom or self-consciousness which authorizes us to have desires which we must then (secondarily, on this way of putting things) endorse in order to make them into actions. So, the threat that the self faces in being confronted by the other is that its ability to authorize its own desires will be taken away from it and only the process of reflective endorsement will remain. The threat is thus quite literally that the self's desires will not be its own, but that, through reflective endorsement, these desires will nonetheless be normative for the self. I suppose that we might call this the slave-zombie view: capable of endorsing desires, but not capable of producing them. This, indeed, sounds like a fate worse than

death.²¹² However, the slave-zombie view is an absurdity and it is the work of the rest of the master-slave section (indeed, the rest of Hegel's social philosophy) to show why this is so.

The Master-Slave Dialectic: Agency and History

Hegel writes that through the battle to the death, the unity of self-consciousness with itself, as exemplified by its two parts, is that: "Through this there is posited a pure self-consciousness, and a consciousness which is not purely for itself but for another, i.e. is a merely *immediate* consciousness, of consciousness in the form of *thinghood*."²¹³ The thinghood to which the first (slave) self has been reduced is now merely a mechanism, no longer operating under its own freedom. It has become the zombie-slave. The master, however, "is the consciousness that exists *for itself*, but no longer merely the Notion of such a consciousness, i.e. through a consciousness whose nature it is to be

²¹² We might also note that the pre-conflict self is not really a full self at all yet because its desires are not mediated by an other. It is merely the possessor of a private language, bound by nothing more than material nature. Such a self has no autonomy because it has no other who will hold it to its concepts or language. The master-slave dialectic is really about becoming a self capable of autonomy.

²¹³ Hegel. *Phenomenology*. 115. "Durch sie ein reines Selbstbewußtseyn, und ein Bewußtseyn gesetzt, welches nicht rein für sich, sondern für ein anderes, das heißt, als *seyendes* Bewußtseyn in der Gestalt der *Dingheit* ist." *Phänomenologie des Geistes*. 110.

bound up with an existence that is independent, or thinghood in general”.²¹⁴ So the master is now a being who exists in radical freedom, having desires but not actually acting on them because another acts for him. Thus the slave has released the master from having to interact with the world.

This leads to an odd sounding situation in which the master has desires and the slave reflectively endorses these desires. The slave is an instrument, a means, to the master’s ends. One might say the master and the slave together make up an autonomous self. But Hegel’s aim is to show the shifting balance between domination and enslavement. The master’s power is to have his desires— any desires— realized by the slave. This is the power of absolute freedom. The slave’s bondage is precisely that he has to make the master’s desire real. But in being so bound, the slave is already experiencing his own autonomy: acting on a freedom that both is and is not his. By obeying the master, the slave is free in the sense that he chooses to follow the master’s orders rather than die. This is the first step in the movement away from the immediate radical conception of freedom toward the mediated, intersubjective conception of freedom Hegel ultimately wants. It is, to be sure, a rather casuistic point, one that philosophy can only make once things have improved and there is more freedom all around.

²¹⁴ Hegel. *Phenomenology*. 115. “ist das für sich seyende Bewußtseyn, aber nicht mehr nur der Begriff desselben, sondern für sich seyendes Bewußtseyn, welches durch ein *anderes* Bewußtseyn mit sich vermittelt [ist].” *Phänomenologie des Geistes*. 110.

Let us turn to the master. The master, Hegel says, dominates the slave because he was prepared to sacrifice himself completely in order to preserve his radical freedom or his self-consciousness, as I have been calling it. He was prepared to sacrifice the very possibility of reflectively endorsing desires, of interacting with the world in concrete terms, in order to maintain his radical freedom. When he finally has the slave under his power, it appears that he has gained absolute power over the world, because he now has someone under his power who will mediate the world for him. All the master has to do is desire—the slave will do his bidding. Hegel writes that the master enjoys his desires “purely”, because he does not have to work to enjoy his desires and the material world thus does not limit him. His desires alone are reasons for the slave to act on the master’s behalf. “What the slave does is really the action of the master.”²¹⁵ The master is pure will but no action.

The master’s will is pure domination and absolute control. But because nothing resists it, it cannot ‘dominate’ anything either. Power, for Hegel, exists in its exercise. Radical or posited freedom is worth nothing without being reflected in action. (This explains why, once the slave realizes that all the power of labor rests really in him, he can just pick up and walk off. The purity of the domination of the slave by the master is that it exists only in the slave’s mind.)

²¹⁵ Hegel. *Phenomenology*. 116. “Was der Knecht tut, ist eigentlich Thun des Herr.”

Phänomenologie des Geistes. 113.

Let us now turn to the slave. I earlier characterized the slave's will as not his own and thus as purely a means to the master's ends. But I also suggested that this slave-zombie view is an impossibility. It is now time to see why Hegel thinks that it is impossible for the slave to be merely (purely) an instrument of the master. This point will also move us from the idea of the preeminence of self-positing freedom to the idea that freedom lies in reflected action.

The transformative event in the slave's life, what makes him a slave, is, according to Hegel, the "pure universal movement, the absolute melting-away of everything stable, [...] the simple, essential nature of self-consciousness, absolute negativity, *pure being-for-self*, which consequently is *implicit* in this consciousness."²¹⁶ This absolute dissolution of the subject's pure or radical freedom, I would like to argue, is what makes the subject intersubjective because it reveals radical freedom to be *for-itself*, subjective rather than just consciousness in general. Radical freedom, I take Hegel to be saying, has a purpose, and this purpose is to become an *an und für sich* rather than merely a *für sich*. This moment of pure being for self, Hegel continues, "is also *explicit* for the bondsman, for in the master it exists for him as his *object*. Furthermore, his consciousness is not this dissolution of everything stable merely in principle, in

²¹⁶ Hegel. *Phenomenology*. 117. "Diese reine allgemeine Bewegung, das absolute flüssigwerden alles Bestehens" which is "das einfache Wesen des Selbstbewußtseyns, die absolute Negativität, das *reine Fürsichseyn*, das hiemit *an* diesem Bewußtseyn ist." *Phänomenologie des Geistes*.

114.

his service he *actually* brings this about.”²¹⁷ Let me gloss this rather difficult passage thus: what is to be overcome through work is the arbitrariness of one’s radically free desires. The way to give radical freedom content is to shape radical freedom through reflective endorsement, engagement with the world of things which, at this point (and in fact, forever) is mediated by the master.

Let me put the point differently. The movement from the quasi epistemological attitude of the subject to an intersubjective attitude is accomplished through the process of reflective endorsement, i.e. where thought meets world and thus where thought has to satisfy itself *in* the world of material things rather than *in* the mind.

This transition, complex as it is, is solved by the development of instrumental reason. The self is merely trying to satisfy its desires. But when the self meets another self— a self which threatens to extinguish the possibility of even having desires at all— it must decide between death and giving up its ability to posit and fulfill its own desires. If the subject chooses the life of servitude, its reasons for action will come from outside it. This means that its radical freedom will be replaced with the radical freedom of another. The slave will retain its ability to reflect on these desires and carry them out. This means that the slave will be acting on the desires of others just as if they were his own. Radical freedom has

²¹⁷ Hegel. *Phenomenology*. 117. “ist ferner nicht nur diese allgemeine Auflösung *überhaupt*, sondern im Dienen vollbringt es sie *wirklich*; es hebt darin in allen *einzelnen* Momenten seine Anhänglichkeit an natürliches Daseyn auf, und arbeitet dasselbe hinweg.” *Phänomenologie des Geistes*. 114.

now given way to a different conception of freedom, one which includes the other as well and which treats the other as the source of constraints on the slave's action. (The whole dialectic of recognition is, taken in this sense, a movement from radical constraint by the other to only relative constraint by the other.)

So, to get back to the point made in the quotation above, the slave's animal nature is no longer the only source of desires for the slave. In working for the master, the slave must now take into consideration both the desires of the master and the technical or material constraints of the world itself. He must bring together not his desires but someone else's objectives with nature in accordance with instrumental reason. What still remains to be seen is how the slave recovers his own subjectivity. This will occur when the slave recognizes that domination by the master is only contingent and that the slave has been exercising his freedom in reflection all along.

In order to understand the move from instrumental reason to intersubjective reason we must understand the paradox present in the slave's situation. On the one side, he takes orders from another such that he fulfills *only* the desires of another and on the other hand, the slave must choose in which way to fulfill these desires. Thus, the slave's radical (subjective) freedom has been replaced with the other's imperative or what, to the slave at least, appears as an objective imperative: the master's words, 'do this or I will kill you'. But what has not been replaced is the moment of agency itself. Indeed, agency now takes center stage as the locus of the slave's 'subjectivity' in that it is only through reflecting the master's desires in action that the slave fulfills the master's desires.

We can now see that the emancipation of the slave occurs when the slave realizes that his 'self' resides not in his radical freedom but in turning desires (the master's or otherwise) into actions: the slave must recognize that the master is *not* the only constraint, but merely one constraint among many.²¹⁸ So what I have been arguing is that domination by the master has already made the slave intersubjective, by revealing that action, not thought alone, constitutes the self. The story of the struggle of recognition is thus the struggle of subjects to replace others' reasons for action with their own. Autonomy, however, is born in action, because action alone is capable of being autonomous in a non-tautologous or public sense. Thought, by contrast, always threatens to be merely private.

In examining the slave's rebirth as a proto-autonomous agent, we must pay particular attention to what the process of coming to understand his autonomy looks like to *him*. Here, again, we should notice that there is a difference in perspective between seeing the possibility of recognition or intersubjectivity from the perspective of the philosopher, from the outside, and that of the suffering agent seeing or not seeing the distant promise of recognition from the inside, from within history.

So let us now turn to the slave's perspective: the perspective of the actor at the beginning of history. This perspective is that of being forced to labor for another. The slave's desires are not his own but the way he carries them out,

²¹⁸ This is clearly analogous to Korsgaard's argument above (ch. 5) that others' desires are fundamentally of the same kind as our own.

endorses them reflectively, is his. What I would like to suggest is that Hegel's theory of agency, in the form of labor, allows us to reconceptualize the interaction between the radical freedom given by self-consciousness and the relative freedom given by reflectively endorsed action. The complete liquification of the self, brought about by defeat in battle, transforms radical freedom from purely subjective into intersubjective freedom because, in defeat, the slave is forced to accept other people's desires as motivating for him.

But what is it about labor that makes this transition possible? Hegel writes: "Work [...] is desire held in check, fleetingness staved off; in other words, work forms and shapes the thing."²¹⁹ Labor is power exerted on to the world in order to transform it. By creating objects, Hegel says, the slave not only comes to appreciate his ability to create but also realizes his negative power of destruction. The fact that the slave has been transformed by existential fear only heightens this point. The existential fear of death quite literally broadens the slave's consciousness to include the desires of the master. The continued fear of death, which we might broaden (in later stages) to include the fear of social death, keeps the subject attuned to the desires of others. The fear of the master means that the slave's reasons for doing what he does stem *from* an other and hence that what he produces is *for* an other. The externality of the product of the slave's work also makes the slave's creations *real* or intersubjective. That is, not only has thinking itself become intersubjective, but so has action.

²¹⁹ Hegel. *Phenomenology*. 118. "Die Arbeit [...] ist *gehemmte* Begierde, *aufgehaltenes* Verschwinden, oder sie *bildet*." *Phänomenologie des Geistes*. 115.

We might tie this to Korsgaard's theory of the constitutive standards of action in the sense that whereas before the confrontation with the other, the slave's actions, say getting food, merely had to conform to some technical standards (like providing enough calories for him to survive), the food he has to provide for the master must now meet some public or intersubjective standards. Hegel writes: "the formative activity has [...] significance [in] that in it the pure being-for-itself of the servile consciousness acquires an existence".²²⁰ Only in having to conform to intersubjective standards is an act ever really an action. And in this way we see that the slave regains some notion of self through his actions. Action itself reflects back onto thought, making it intersubjective too.

Now, if the slave realizes that his labor, alienated though it is, contains irreducible elements of his own thinking and execution, the slave will realize that his labor is, in a constitutive sense, what makes him free. The slave is free not only in the sense that he can choose death but also in the sense that he must interpret the master's desires, making his own imprint on the final product. Alienation, in other words, can only be experienced by a subject.

To sum up the movement from self-consciousness to intersubjectivity, we can now say that the slave has lost his radical freedom but has gained the capacity to see others' ends as reasons for him to act. This change is itself the transition from consciousness to self-consciousness or intersubjectivity. The enslavement to the will of the other has now emerged as a constraint on the

²²⁰ Hegel. *Phenomenology*. 118. "im Bilden wird das Fürsichseyn als *sein eigenes* für es, und es kömmt zum Bewußtseyn, daß es selbst an und für sich ist." *Phänomenologie des Geistes*. 115.

slave's action in the form of the requirement to conform to intersubjective standards of action. The slave only develops a self when he realizes that taking into consideration the desires of others allows a richer life than the desire-based existence of the master. So, we can say, the seeds for a theory of recognition and with it for a theory of justice, have been sown.

Attaining intersubjectivity means that the slave's immediate desires are replaced by an ever expanding set of reasons. This then means that instead of endorsing only his own desires, the slave ends up endorsing desires which come from outside his immediate purview, and this means that his reasons are now intersubjective instead of subjective. What I am trying to argue is that the dialectic of the master and slave replaces radical freedom with a concrete process of reason giving, or reflective endorsement, and that this shows that reason giving can *only* be construed as the ability to endorse this or that norm here and now. For Hegel, then, normativity is necessarily a process of the reflective endorsement of those reasons presently available to the agent. There is no norm that holds for now and all times but only what counts at the moment. This, too, is Williams' point.

However, lest we fear that Hegel is falling into a relativism, we must also note the second element in this process, namely self-consciousness. For, as we have seen, self-consciousness construed as radical freedom has been replaced by self-consciousness construed as the freedom of all. Self-consciousness is now no longer merely the immediate consciousness of one's own desires but rather the consciousness of the desires of all. Reflective endorsement no longer

selects its action merely from its immediate set of desires but from the universal pool of the desires of all furnished by self-consciousness.

Conceiving of self-consciousness in this way allows us to see the importance of Hegel's notion of Geist. The development of Geist is precisely the coming into actuality of the possibility of taking the desires of others to be one's own reasons. The dialectic of recognition thus takes place between the fundamental capacity of self-consciousness (in each individual) to take others' desires as reasons for action and the limited capability of reflective endorsement to turn the consideration of others into a concrete reality.

And this, too, is where history enters into the picture. History, for Hegel, is the process in which agents act on the possibility of recognition, on the possibility of taking other people's desires as reasons. But history is also the continued failure of this ideal and so keeps reminding us of the fact that we have not yet reached the point of full intersubjectivity. So, to bring things back to the master-slave dialectic, we can say that the master-slave dialectic is the origin of history in the sense that the dialectic is also the origin of normativity proper. For in acting on the command of the master, the slave submits himself to a will which is not his own, but then realizes that he is bound not by the master's physical power but by other people's desires cum reasons. But other people's desires also include his own desires, so the slave has now arrived at some notion of mediated or socially constrained agency. The master, it is important to note, is left behind because he does not experience or enter into the normativity of the world around him; he

does not bind himself to any goals because all of his desires are satisfied by the slave. The master thus remains outside history.

I would thus like to interpret the ostensible inequality which marks the origin of normativity as necessary for this development itself. It is only in inequality, in suffering or limitation, that normativity appears as such at all. Normativity enters the picture because not everything is in its place, because the world does not yield to our mere thoughts (as it does, we may imagine, for the master).

By way of concluding this section, I would like to say that Hegel's presentation of the master-slave dialectic, as I have interpreted it, shows that normativity is only ever experienced from the perspective of the loser of the conflict. The winner, like the master, never enters normativity because he is never at odds with existing conditions. Because he does not experience suffering, and thus has nothing to hope for, he remains outside history. The loser, by contrast, exists in conflict and continues to insist that his desires be recognized by others. The difference between these two perspectives is quite nicely captured in Hegel's elucidation of the *Doppelsatz*, according to which Hegel's dictum that the real is the rational really means *the real must become the rational*.²²¹ For we can now see that to think that the real *is* the rational is the

²²¹ GPR. Vorrede. Hegel writes: "Das Vernünftige ist das Wirkliche; und das Wirkliche ist das Vernünftige". On the elucidation given by Hegel to Heine see Hardimon. *Hegel's Practical Philosophy*.

perspective of the master who exists outside of history. The slave's perspective, by contrast, is the imperative that the real *must become* rational. The normative is here characterized negatively, as the failure of the present order of things. The failure of the present order means that some injustice exists and that we must strive for justice.

Pinkard's Categorical Reading and the Perspective of the Slave

The above interpretation of the slave's perspective sheds some light on the problem of the transcendental interpretation of Hegel's thought by Pippin. Pippin's interpretation was characterized by a strong insistence that the categories of thought could be reconstructed or constructed from within self-consciousness itself. This means on the one hand that all ways of understanding the world are, in principle, available already and, on the other hand, that changes in the world will not affect our conceptual schemes. Pippin rightly rejected this set of claims as both impossible to prove and implausible. The problem with this strong interpretation of Hegel's theory is that it appears to be entirely written from the perspective of the master, that is, from the perspective of one who has already worked out all of the problems, metaphysical and physical, which might face him. That is, to one who has already worked out the conceptual scheme of the ages, the real already *is* the rational and the rational already *is* the real.

But this is not, I have been arguing, the right way to understand Hegel's work. So, given the failure of Pippin's forceful and ambitious interpretation of Hegel, it seems worthwhile to consider whether it might not be possible to

provide a more modest version of Hegel's speculative philosophy, a version which could be told from the perspective of the slave, the agent *within* history. Such a theory could avoid the claim of Pippin's transcendental thesis to the effect that Hegel must give an account of the totality of the conceptual scheme *per se*. Hegel's speculative philosophy might rather be understood as working out a determinate view of the *human* and temporally bound conceptual schemata. This would open up the possibility of a historical revision of Hegel's view, but retain Hegel's apperception thesis about the development of concepts. Such a view would interpret conceptual development as historical and thus open-ended.

Terry Pinkard's interpretation of Hegel does just that. It allows us to see the conditions for the possibility of the theory of recognition as I have developed it in the above discussion of the master-slave dialectic. I propose now to take up this other perspective on the problem of self-consciousness in Hegel. In concluding this discussion, I will say a little more about the possibility of justice made possible by Pinkard's interpretation.

Pinkard's work on Hegel has shadowed Pippin's work over the past 30 years and both writers share many of the same aims in interpreting Hegel.²²² But Pinkard's reading differs from Pippin's reading in that Pinkard thinks that Hegel

²²² Both wish to show that Hegel is a non- or post-metaphysical writer, both agree that he owes a great debt to Kant and both think of Hegel as a radical social critic. Pinkard. "How Kantian Was Hegel?".

can be usefully read as a category theorist.²²³ I will explore this alternative in what follows. The aims of this alternative will be to see whether there can be a non-metaphysical reading of Hegel which can say something more about the self-grounding of reason in Hegel than either Pippin or Brandom could. For Pinkard, the dialectic may not give us the transcendental conditions of knowledge, but it can tell us if thought is consistent and coherent.²²⁴

Let us turn first to Pinkard's criticism of Pippin's approach in *Hegel's Idealism*.²²⁵ This will give us a better understanding of what the categorial reading can do in grounding Hegel's notion of reason. Pinkard agrees with Pippin that the place of the apperception thesis is most evident is in the *Phenomenology*, since the *Phenomenology* is supposed to show that all the so-called metaphysically transcendental objects of knowledge may in fact be regarded as posits of thought. But, Pinkard contends, even if the *Phenomenology* were successful in its own terms, it would not show that thought must have such and such a categorial structure; it would only show that it would be possible to construct such a pure categorial structure of thought. But Pippin has denied that

²²³ For an earlier and influential view of what such a reading might look like, see Klaus Hartmann. "Hegel; A Non-Metaphysical View".

²²⁴ Pinkard. "How Kantian Was Hegel?" *Review of Metaphysics*. 43 (1990) : 831-38. 837

²²⁵ I will take Pinkard's critique of Pippin's transcendental view to be relevant to all approaches which do not pay sufficient attention to the perspective of the agent within the historical fray. This

Hegel succeeds in proving that the categories of thought he develops are developed through determinate negation. Hegel's categories are thus not thoroughly rational.

The categorial reading, however, allows Hegel a way out of these damaging charges: on the categorial view, Hegel is a speculative philosopher, and not a transcendental philosopher. Speculation is an alternative to deduction, in the logical sense. While in a deduction nothing new is added (the conclusion is already contained in the premise and needs only be worked out), in speculation, something new is *posited*.²²⁶ According to Pinkard:

Hegel's goal is *not* to show the transcendental conditions without which there can be no thought at all, or the conditions under which any conceptual framework is possible. It is rather to explain the rational possibility of certain categories. The assumption is that contradictory sets of categories are impossible (after all, the real is the rational), and one explains how a category is possible if one explains how both it and its apparent contradictory category are each compatible with the other.²²⁷

While Pinkard agrees with Pippin that the transcendental thesis according to which the world's conceptual scheme can be worked out for all time fails, the speculative philosopher still stands a chance of providing some sort of a grounding of reason, though not one as determinate as Pippin thought it must be. We can thus take Hegel's dialectical philosophy to rest on the more modest thesis that there are resolutions to *our* problems of categorial possibility and that

criticism thus also touches Rawls and Korsgaard's arguments for the conditions of the possibility of justice.

²²⁶ Pinkard. "The Categorial Satisfaction of Self-Reflexive Reason". 9-10.

there will be new problems and new resolutions that we cannot foresee. This allows Hegel to be open to new conceptual developments in history.²²⁸ Such an approach also transforms the apperception thesis into a thesis about *our* conceptual possibility rather than about conceptual possibility in general, and thus adds a genetic approach to conceptualization to the theory.

From Metaphysics to Ontology

As with Pippin, the real crux of the argument Pinkard advances is to be found not so much in the description of the social categories which Hegel is taken to be describing as in the speculative justification for such description. In examining Pinkard's proposed reshaping of the apperception thesis, we must pay particular attention to how the transcendental approach is replaced by the categorial approach. In other words, what would it mean if conceptualization were rational, at least as it pertains to our situated and contingent human selves?

It should be noted that I will be examining the categorial interpretation as an alternative to the transcendental interpretation offered by Pippin rather than in its own right as a defense of Hegel. What I am primarily interested in is what sort of resources the categorial view can provide us with in our attempt to ground rationality immanently. I cannot, here, examine the difficult issue of what, exactly, for Pinkard, consistency and negation mean. That is, we will have to defer the

²²⁷ Pinkard. "The Categorial Satisfaction of Self-Reflexive Reason". 11.

²²⁸ Pinkard. "The Categorial Satisfaction of Self-Reflexive Reason". 11.

question of whether Pinkard's views about Hegel's speculative philosophy are correct and attend only to what promises such a view holds for several of the issues which have been raised throughout this dissertation; for instance, the importance of history and the significance of agency in such a contingent context.

Like Pippin, Pinkard believes that the key to understanding Hegel's speculative philosophy comes from an analysis of the *Logic*. Here the Logic of Essence is of particular importance.²²⁹ In the Logic of Essence, Hegel diagnoses the two conflicting metaphysical claims of his predecessors (the rationalist and the Kantian) in the following way. The realist view posits an essence behind appearance which is itself beyond appearance and thus deprives itself of a way of knowing the ground of appearance. *Objective realism* thus results in skepticism. The other formulation of essence, to avoid positing a substance behind appearance by saying that it is the mind itself which imposes the basic categorial determinations of our understanding, however, lands us in Kantian *subjective idealism*. Here we lose the world.

The answer to this dilemma is to be found in Hegel's doctrine of the concept which Pinkard interprets as the study of inferences, both formal and material. Appearance is determinate in the way it is simply because it is the way we must think it, the way in which the world necessarily manifests itself to us. A concept is thus a *position* in a scheme of inferences and has no metaphysical

²²⁹ Pinkard's interpretation of Hegel's *Logic* is at least congruent with Pippin's in the detailed discussion of particular passages. What the two differ on is to what standards Hegel's claims about his goals and achievements should be held.

substantiality; outside of such a scheme of inference, a concept has no existence whatsoever.²³⁰

What is arrived at in Hegel's doctrine of the concept is Absolute Idealism which is (in Pinkard's summary): "the doctrine (1) that there is no metaphysical intermediary that either stands between our concepts and the world; or (2) that we do not need to posit a metaphysical entity that stands 'behind' (or 'beyond') appearance to explain its determinateness."²³¹

The determinateness of appearance can be perfectly well explained in terms of systematic complexes of concepts, and concepts themselves can be explained by the inferential roles they play *vis-à-vis* each other. Since the world manifests itself to us in systems of concepts, any general skepticism about the relation of knowledge and the world is misplaced. Pinkard says that the Doctrine of the Concept, the last section of the *Logic*, replaces metaphysics with ontology or, as we have been saying, speculation. We still can postulate entities for the purpose of theory construction (like genes, or leptons) but we need not think that we need to postulate *metaphysical* entities to do the kind of philosophical explanatory work that we earlier thought necessary.²³²

Conceptual thought is thus not only capable of giving an account of the world but also of giving an account of itself. It is self-subsuming and self-explanatory in the sense that the determinateness of concepts is explained by

²³⁰ Pinkard. "The Successor to Metaphysics". 304.

²³¹ Pinkard. "The Successor to Metaphysics". 305.

²³² Pinkard. "The Successor to Metaphysics". 305.

the kind of inferential moves made with the set of concepts itself. The theory is absolute in the sense that there is nothing outside it and in the sense that there is no alternative to displace it. There is no meta-concept. The categories of Absolute Idealism express the way the world is. We should note here that the claim that Hegel's theory subsumes itself under itself is a key claim for Pinkard, for it states that there can be no outside to theory building. This claim is meant to oppose Hegel's tendency, as noted by Pippin, to make claims about the completeness of his system; for the claim to completeness seems to entail an outside which is *not* subsumable into the system. This is the metaphysical remnant which Pinkard seeks to avoid. More so than Pippin, Pinkard is thus not shy of using Hegel to overcome Hegel.

Before we move on to a discussion of Pinkard's interpretation of Hegel's social theory, it is worth noting that, while Pinkard follows the movement of Hegel's system from Logic (skipping the philosophy of nature) to *Realphilosophie* or Philosophy of Spirit, Pippin's book proceeds in the opposite direction. The point is significant because it makes a difference to our understanding of the two theses in Pippin's book which way we proceed. What is at stake here is whether Pippin's transcendental thesis— that the categories of reason require a determinate logic— is also a requirement of the human understanding. At one level, of course, Pippin is right to demand this, as Hegel's anti-realism entails that the categories of reason can only ever be *for us*. Pinkard's view, however, makes room for the possibility that we have not yet worked out what the necessary

categories of reason are (as they truly are) and shows that we must content ourselves with trying to work out the coherence of *our* categories of reason as they appear to *us* from *our particular* place in history.

Spirit's Purposiveness

Jettisoning the transcendental thesis renews the question of what a self-grounding rational form of life might look like. Part of the value of the categorial reading of Hegel was to suggest that our theory of *the* world contains within it a theory of *our* theory of the world, i.e. a way of understanding our own theorization. This means we must have some notion of our own place in history. Pinkard argues that this notion of our place in history can be conceived of in terms of purposiveness. Purposiveness, however, is not teleology since the latter implies a driving force which stands outside of the process while the former does not. Pinkard contends that it is part of the *Logic's* solution to the problem of ground that we take ourselves as trying to give an ever more complete account of the relation between different parts of our conceptual world. What Hegel seems to be after, Pinkard contends, is that a set of inferential connections can 'point to' its rational successor.²³³

Hegel calls this constellation of inferential relations which each seek to become more rational, Spirit. Spirit is also Hegel's name for the process of developing concrete social and political freedom. Spirit is absolute for Hegel

²³³ Pinkard. "The Successor to Metaphysics". 316.

because it is not mediated by anything outside of itself and must provide its own necessity. This relation, however, is more than a mere coherentism because the dialectical movement relies on its own internal dynamism brought forth by historical contradiction in the form of suffering. The goal of Hegelian ethical theory is thus to seek a workable conception of the principles of personal life and cooperative social life without having to invoke any metaphysical conception of the person or society. This is not done by some 'overlapping consensus' as in Rawls. Rawls is merely agnostic about metaphysics while Hegel rejects it. So Rawls can only offer us a theory of what we moderns would rationally choose from an appropriate menu, but he cannot explain how it is that such and such things are on the menu and why we are the people for whom such things seem like options.²³⁴

Dialectic, Negativity and Necessity

The answer to the question about the availability of historical alternatives is both historical and philosophical. To answer it, we will also need to know more about how to distinguish Hegel's notion of reason from Rawls' notion of rational deliberation. This will help us see how Hegel does indeed reject metaphysics of both the Kantian and realist varieties. This point is all the more pressing, since, as noted, certain of both Pippin's and Pinkard's interpretations of Hegel do not seem very different from Rawls' notion of social cohesion. Part of the argument

²³⁴ Pinkard. "Self-understanding and Self-Realizing Spirit in Hegel's Ethical theory". 89.

Pinkard seems to want to make, however, is that the one central difference between them, the question of history, is not merely one difference among many, but the reason why Hegel's account of *Sittlichkeit* has a rational speculative grounding while Rawls' account of justice does not.

Pinkard argues that Hegel's central philosophical concern is for the self-authentication of our inferences. Given that there is no external authority, this means that all authority must be immanent to the authorizing process. Such internality of authority means that all theoretical accounts of reason must be examined in terms of their ability to generate and justify inferences or reflection. For Pinkard, justification means that individuals have to determine whether their position in 'social space' licenses their intended actions.²³⁵ The validity of the inference can be determined by whether the inference and the account support one another. If the inference is not valid, what Hegel calls negativity arises.

In order for this negativity to carry any weight, Pinkard adds a second thesis to that of internal consistency: the ground rules of the social space, the rules for licensing inferences, appear to the people in the social space not just as 'given' or optional but as necessary. The ground rules appear as certain and also as structuring what counts as truth. Part of a theory of knowledge, therefore, must be to determine if there is any set of authoritative reasons that can generate their own necessity in a way that does not undermine itself. Understanding this necessity means becoming self-conscious about norms and becoming self-conscious about the apparent paradoxes, incoherencies and conflict within our

²³⁵ Pinkard. *Hegel Phenomenology*. 7.

norms.²³⁶ The dialectic is constituted by the appearance of these internal inconsistencies.

I have tried to give an illustration of such an account in the discussion of the master-slave dialectic above. There I argued that the achievement of recognition or intersubjectivity lay in the slave's realization that the constraint placed on him by the master was just one constraint among many, and that such a constraint was, in fact, really a constraint that he placed upon himself. To recognize that one is autonomous despite the prevailing coercive norms is to have reason or be authorized by oneself to seek to change the current situation. Inconsistencies in the structure of reality are normative for the subject who exists within it.

The question of the structure of (current) reality brings us back to the fundamental question of the determinate schema of reason. Recall that Pippin argued that the dialectic, in Hegel's account of it at least, does not give a satisfactory account of the determinacy of reason because it does not furnish us with satisfactory categories according to which we can measure and evaluate our concepts and actions. The only process strong enough to provide an authoritative reconstruction of thought, determinate negation, is not borne out by Hegel's writings. We have been looking for an alternative account to such determinate negation in Pinkard's account and this is the point at which we must evaluate Pinkard's final theory.

²³⁶ Pinkard. *Hegel Phenomenology*. 7.

From Metaphysical Truth to Social Truth

The answer to how the dialectic is supposed to work, Pinkard contends, is to be found in Hegel's (modified) apperception thesis itself, and not in the transcendental thesis, as Pippin argues. Let me recall that the recreation of Pinkard's account of the self-authorization of reason is only a sketch and not an actual defense. It is nonetheless useful to provide some sort of detail so that we may better see the consequences of this view, should it prove to be correct.

Pinkard thinks that normativity is provided by the apperception thesis itself. For Pinkard, as for Pippin, the apperception thesis states that there is no given which we must assimilate or receive in order to know the truth. Rather, what the apperception thesis means is that subjectivity must proceed from the unity of concept and object rather than from their separation. And this means, Pinkard argues, approvingly citing Brandom, that the distinction between the normative and the factual is itself normative.²³⁷

Pinkard takes this argument to mean that, for Hegel, nature itself, as mere fact, is dead and can only be animated by a subject constructing itself. He argues that, since nature is finite and finite things have their authority outside of themselves, nature cannot tell us anything about who we actually are. The consequence of the apperception thesis, however, is that subjects, as self-authorizing entities, are infinite. As self-authorizing, subjects construct their own meaning and this meaning is what Hegel calls reality. Thus reality, for Hegel, is

²³⁷ Pinkard. "Innen, Außen und Lebensformen: Hegel und Wittgenstein". 280.

no longer a question of fit between conceptions of subject and the world. In reorienting the question of truth, Hegel also rejects coherentism, realism and other forms of truth seeking where truth is conceived as the relation between the world and the mind. For Hegel, the only relation that can produce truth is that between subjects.

In terms of the social practices which we engage in to construct truth, this means:

We are always already 'inside' as well as 'outside' of ourselves in that, as self-interpreting animals, we are 'absolute negativity'; we are only ever concerned about whether we correspond to our conception of ourselves, and whether the activities which make up our lives in our institutions, are true, or even if the self which we are supposed to be faithful to (according to the modern ideal of authenticity) is really our true self. ²³⁸

The historical authenticity and the refutation of the charge of historical relativism is developed in Hegel's theory of the negativity of history. For, as Pinkard notes, normativity is normative only if it at once seeks to generalize, give people reasons for acting, and also, sometimes, fails to give them reasons. The dialectic of reason giving can best be seen in historical development organized around the dissolution of norms. The most famous example of this is Hegel's

²³⁸ Pinkard. "Innen, Außen und Lebensformen: Hegel und Wittgenstein". 283-84. My translation. "Wir sind immer sowohl ‚innerhalb‘ als auch ‚außerhalb‘ unserer selbst, dadurch dass wir, als selbstinterpretierende Tiere, ‚absolute Negativität‘ sind; wir sind immer darum besorgt, ob wir unserem ‚Begriff‘ von uns selbst entsprechen und ob die Tätigkeiten, die unser Leben in unseren Institutionen ausfüllen, wahr sind oder sogar ob das Selbst, dem wir treu sein sollen (gemäß dem modernen Ideal der Authentizität), selbst unser ‚wirkliches‘, wahres Selbst ist."

discussion of Antigone, whose actions are the result of the deep commitment to her society but also reveal the limitation of the practices of the historical Greek city state. The fact that Antigone can both uphold the norms of her time and break them shows that contemporary normativity is only the condition for our actions rather than the reason for our action.²³⁹ To put this slightly differently, we might say that present norms are strong reasons for us to act in a certain way but not a determining reason for us to do so. We also have our own set of desires, at odds with the current norms, which also make a claim on us. As the example of Antigone shows, sometime to act on our own desires, on our conscience, as it were, is to help establish a more adequate form for social norms. (This, of course, is the dialectical version of Williams' insistence on the irreducibility of subjective desires.)

Hegel reworks Kant's principles of ethics in a more substantive way.²⁴⁰ There is rational deliberation about action, just as in Kant, but the principles according to which actions are evaluated are the substantive commitments of one's society. The dissolution of a form of life, Hegel argues, is a necessary part of historical progress to the most rational and complete understanding of human society, the self-understanding of Spirit.

²³⁹ Pinkard. "Innen, Außen und Lebensformen: Hegel und Wittgenstein". 286.

²⁴⁰ Pinkard. "Inne, Außen und Lebensformen: Hegel und Wittgenstein". 290.

Revisability: Our History, Our Concepts

Pinkard thus seems to agree with my claim that the genetic account of the development of human subjectivity given by the categorial interpretation does not supplement Rawlsian coherentism with a historical perspective, but that history is itself the product of a struggle for conceptual coherence. This means that human rationality has two sides; the strong demand for the rational coherence of the conceptual scheme on the one hand, and the demand for the conceptual revisability of such a scheme on the other hand. What makes Rawlsian coherentism possible, what, in other words, makes it possible for a reflexive equilibrium to come about is the very historical situatedness of the agents debating and their ability to redefine themselves in the act of deliberation. This scenario can, again, quite usefully be illustrated by Harris' view that freedom is at once a radical positing of freedom and also a free acceptance of the contingent determinations that such positing gives rise to.

Our norms are normative for us because they are historically given as both contingent and necessary. They are necessary as the norms that we must accept as subjects living at a certain place and time, but become contingent to the extent to which we are able to posit our way out of them by working toward a new social arrangement which is more coherent and thus conceived of as necessary.

Revisability is thus central to Pinkard's categorial interpretation. As we saw, a central difference between the transcendental thesis and the categorial interpretation is that on the categorial reading Hegel need not be understood as

having developed a determinate conceptual structure for this and for all time. By arguing that his philosophy provides the final word on the condition for the possibility of knowledge, the formulation of what such a condition might be threatened to remain outside the system itself and become reified into a metaphysical relic, destabilizing the whole system. Pinkard's emphasis on revisability, by contrast, ensures that the very act of theory building, of developing new conceptions of knowledge, always stays within and is subsumable under the positing of the new theory.²⁴¹

This leads us to a further point of divergence between the transcendental thesis and the categorial reading. The fact that the categorial reading subsumes its own theoretical positing into itself means that Hegel's philosophy is absolute in the sense that there is nothing outside it. But this does not mean that the theory is complete. The claim to completeness that Pippin emphasizes actually obstructs what is most valuable about Hegel's theory: the centrality of dialectical revision of norms brought on by the concrete recognition of suffering. There is room in the categorial reading for advances in natural science as well as in logic.²⁴² More importantly, there is room for a more fully developed theory of justice.

The notion that Hegel's system aims at the absolute rather than at the complete thus sits rather well with the concern for justice of which Pippin writes.

²⁴¹ Pinkard. "How Kantian Was Hegel". 836-37.

²⁴² Pinkard. "How Kantian Was Hegel". 836. See also Hartman. "Hegel; a Non-Metaphysical View". 104.

For it is only through total commitment to the rational examination of all of our concepts that we can hope to advance to a more just society, a society in which all conceptual discrepancies are subject to revision and hence improvement. Indeed, one could go so far as to say that the 'absoluteness' of Hegel's theory is a demand to examine and reevaluate all concepts and social structures. In this context, Hegel's critique of Kant according to which the categorical imperative is a conceptual reification receives new meaning.

Conclusion

In this chapter I have tried to show a way of conceiving of normativity as an activity within a necessarily embodied world. The point has been to show that the possibility of freedom, an idea shared by all writers in this study, is not sufficiently specified by the claim that intersubjectivity is possible. What makes intersubjectivity or recognition possible is that it appears as a demand on us from within history itself.

This demand, I have been arguing, comes to us in the form of injustice. Injustice is the perspective of the agent caught up in a system of norms with which she does not wholly agree or, to put it less cognitively, which subjugates her even without her fully being aware of it. The problem of ideology is a difficult one and far be it from me to take it up here. Suffice it to say that Hegel presents it as a historical imperative and hence also as an achievement to become conscious of one's experience of injustice and, having become conscious of it, to

act on it. I have tried to give an account of the ur-form of this dialectic in my reading of the master-slave dialectic.

The second element of the problematic of justice and philosophical perspective that I have sought to emphasize is the need to understand reflection itself as subject to revision. I have tried to show this in my reading of Pinkard's work. Pinkard's interpretation of Hegel as essentially a theorist of the possible categories of social interaction means that, for Pinkard, Hegel's speculative philosophy always remains open to revision. Revisability, I have then argued, means the possibility of accounting for the perspective of the slave who demands that current norms be changed to accommodate his desires.

I would like to end this chapter by returning to the thought that I have raised throughout this study, namely the overcoming of any distinction between nature and subject or between the moral and the non-moral. The account of the development of normativity in the master-slave dialectic and the account of speculative reason itself, I would like to now say, mean that normativity goes all the way down. Normativity pervades us from the most basic labor of self-preservation to the most difficult moral choice. This is so because the dual processes of self-consciousness as self-positing radical freedom and reflective endorsement of such posited desires are with us no matter what we do. In this sense, there is no distinction between the moral and the non-moral. Whether we take the hardness of a nut as a reason to bash it open with a rock or a suffering person's cries to be a reason to help him or her, we have within us the constraints of our commitments, that is, our other desires. But all constraints are

normative for us. The strength of these commitments, of course, is mediated by the norms which constitute us.

CONCLUSION

This dissertation began as an investigation into the possibility of a philosophical deduction of the categories of morality in Kant's *Groundwork*. The ultimate aim of this deduction for Kant was to find a way of carving out a space for human freedom or agency in the midst of the seemingly overpowering forces of nature. I concluded in the first chapter that it was impossible to deduce the notion of pure practical reason from the idea of freedom. This failure, as one might characterize it, is the failure of reflection to orient our actions with regard to a supreme notion of morality. The problem, to put it another way, is that in order to create a space for freedom Kant's notion of moral reflection opens up a chasm between the intentions of the free will and the determined material world. Morality lies on the one side, nature on the other. This study's subsequent chapters have thus been united by the quest to find, while remaining within the process of reflection, a way of developing a notion of agency that is both material and moral in the sense that it synthesizes the activity and passivity of the human.

This investigation proceeded in three stages corresponding roughly to the three thinkers examined, Rawls, Korsgaard and Hegel. The first strategy of overcoming the gap between nature and the free will is Rawls' attempt to reject the problem entirely by just assuming that agents simply *are* rational. This generated the problem that either Rawls' was assuming more even than rationality (in which case he was assuming what is fundamentally at stake: the principle of pure practical reason) or that the more modest notion of reason he

was assuming was not powerful enough to give rise to the intersubjectivity on which his theory of justice relies. This latter objection takes the form of the thesis that reason is egoistic in the sense that though my thoughts might be normative for me because they are *mine*, and the same goes for you, this does not make *your* thoughts and goals normative for *me*.

Korsgaard's work takes up many of the important questions left unanswered by both Kant and Rawls and develops an immanent account of normativity by arguing that the categorical imperative is normative for us because it has the structure of thinking itself. Thinking or reflection, in Korsgaard, is able to synthesize intentions and the world because it is the condition for the formulation of any action whatsoever. In other words, I must turn my desires into a plan of action and this plan is normative for me because I am its author. This is the basic structure of autonomy. The categorical imperative, on this reading, is not some further test to be applied to a plan of action, it *is* the plan of action itself.

In order to show the last step, Korsgaard also has to address the egoism argument advanced against Rawls. Korsgaard argues that the categorical imperative is not merely a technical constraint on action but an ethical constraint as well. She argues that reason is by its very nature public, because the process of reflective endorsement does not distinguish between internal desires and external reasons in developing a plan of action. Thus your needs and my own needs are on the same level with regard to becoming candidates for my reflective endorsement. Korsgaard, I argued, is thus able to show that intersubjectivity is possible in just the way Rawls thought it was. We are able to deliberate together

and the results of this communal deliberation, this inter-action, are normative for us because we *can* take each others reasons as reasons.

A consequence of Korsgaard's view, I also argued, is that the division between the moral and the non-moral or the natural is eroded to a great extent. Because reflection concerns all actions equally and because other people's reasons enter into our reflection in just the same way (though they do not necessarily have the same value) that considerations about the empirical world do, normativity is principally the same all the way down. This aspect of Korsgaard's thought can be seen in her theory of the constitutive standards of action, which I interpreted as an account of the institutionally mediated way we take other people's reasons as reasons for us. The way in which we take other people's reasons to be our reasons, the way intersubjectivity evolves, is, I argued, historical.

Throughout the discussion of Korsgaard's work, I have also been pressing Williams' objection to the current morality system to the effect that it is a matter of moral luck whether the way we take up other peoples' reasons fits with what is expected of us from society. I take this objection to be the perspective of injustice, of the oppression of one social group by an other. Because Korsgaard's notion of justice contains only a formal demand that we be consistent in our dealings with others, I argued that the power dynamics present in inter-action cannot be accounted for by her. Korsgaard's account of justice, I argued, was incomplete because it cannot account for the perspective of the marginalized

member of the society. If reflection treats everything equally, then differentials of power become invisible. Reflection must have a perspective.

By turning to Hegel's work, I try to present an even more immanent account of reflection, one that would allow us to take the perspective of the dissatisfied member of the society into consideration. Such a theory would be immanent in the sense that it would recognize the demand of the unreconciled individual as normative for its reflection. Such a theory would thus tether the whole process of reflection to the existence of injustice within its plan of action and would continue to adjust itself accordingly.

In order to develop this idea further, I offer an interpretation of the master-slave dialectic in Hegel's *Phenomenology* in which I argue that Hegel thinks a theory of justice can only be proffered if it is also tied to the perspective of the slave, since it is only in the slave that the hope of justice appears. The perspective of the master, the perspective of the one who is satisfied with the current social arrangement, is irrelevant to the question of justice because the master's conceptual scheme has already reached an adequate interpretation of the world. History, I thus argue, is the continued insistence of the slaves to be accommodated within the dominant practices of society. Such a historical perspective is absent in both the work of Rawls and Kant and in the work of Hegel interpreters like Pippin.

I conclude the study by emphasizing the need to conceive of reflection as always occurring from the standpoint of those within the process of history. For it is only in emphasizing this standpoint that we can also make sense of the need

to improve and revise the existing social system. This, I finally argue, is where normativity and autonomy reside.

It is now time to summarize several of the conclusions reached throughout the dissertation. As was indicated in chapter three and fleshed out in the last two chapters, a central concern of this dissertation has been the question of the place of suffering and nature in moral theory. In Kant, this concerns comes in the form of the notion of agency or freedom which tries to move us out of our passivity vis-à-vis nature. To move us away from nature or rather to carve out a space of freedom within nature, however, Kant has to make an argument that is invalid (the deduction of the principle of pure practical reason). But the consequence of Korsgaard's reformulation of Kant's original project in terms of a more immanent understanding of reflection is that there remains, ultimately, no difference between the human and the non-human, since everything is reflected on in just the same way. I do, however, think that Korsgaard's approach is consistent with much of Kant who does seem to think that at some level we are just passive in the face of the pressures other people put on us and that other people are therefore just an extension of physical nature. (In the life and death struggle, it makes no difference whether you get run over by a bus or crushed by a falling tree.)

The problem for this view comes in the way it assimilates all action to ethical action. This can be seen in Rawls and Korsgaard's fine refutation of egoism to the effect that egoism is not an alternative to practical reason since it

too must rely on practical reason to formulate action. Egoism is thus shown to be secondary, a type of behavior and thus not an opposing view. However, in rejecting egoism, Rawls and Korsgaard seem also to be rejecting nature itself in the sense that they view all action as fundamentally able to transcend the contingencies of nature through reflection. But in doing so, they seem to me to be missing what was so intractable about nature in the first place— that it makes us do things that we don't seem to ought to want to do. This is the objection I take Hegel (and Williams) to be raising— the pressures of nature must always be considered in any kind of ethical theory because ethics is something to be achieved (must be fought for against our natures, both inner and outer) rather than a fact about us.

Let us look at this problem a little more closely. The first issue to be considered is the status of Rawls' theory of justice. In chapter two I concluded that Rawls' was unable to generate true communal deliberation because he took for granted the notion of intersubjectivity that he argued for in his work. I would now like to amend my judgment on Rawls in light of Korsgaard's argument for the publicity of reason. I think that Korsgaard's argument for the publicity of reason decisively refutes the egoism charge against her own work and also against Rawls. That is, Rawls' strategy of assuming that my reasons can enter into your considerations without trouble (at least in principle) is borne out by Korsgaard's arguments. Supplemented in this way, I think that Rawls' assumption about the possibility of the reflective equilibrium is sound: it is in principle possible to deliberate together.

To stay with Rawls' notion of the reflective equilibrium for the time being, we could now say that the possibility of deliberation is not its actuality and that, in order to turn deliberation into an actuality, we must also consider the actual perspectives of those deliberating. This is what the Hegelian critique of the Kantian enterprise insists on, as I have been interpreting it. Justice is visible not in theory itself, but rather in the concrete struggle which takes place, sometimes just under the nose of theory. If theory is to take account of justice, it must be able to contain both considerations about the possibility of justice in general and also pay close attention to concrete political and historical activity. Though, as O'Neill correctly points out, no theory can do without abstraction. A theory which idealizes away the insistence of injustice in order to present a more coherent view of 'justice' does more harm than good.²⁴³

To put this point together with the question of suffering, we can see that Rawls' rejection of metaphysics, that is, with the question about agency in the face of nature itself, really means that there is no way for his theory to properly interface with the material world. As suggested earlier, Rawls' and also Korsgaard's theories do not register the pressures of the material world (both physical and social) that occur in each agent's attempt to be her rational self. In other words, the struggle to become rational only really makes sense as a

²⁴³ Onora O'Neill. "Ethical Reasoning and Ideological Pluralism". I am generally in agreement with O'Neill that Rawls is not guilty of much idealization, but there is a sense in which abstraction becomes idealization when it is not done at the right level.

response to suffering (that is, nature). And this is precisely what Rawls has cut out of his theory.

This question feeds into the more general question that has been in the background of this study all along, namely the question of the overcoming of metaphysics and its replacement with something like speculative philosophy, or a philosophy of immanence. While it is clear that all writers I have been considering explicitly reject the label of metaphysician, there has been considerable question about unresolved metaphysical commitments even in these anti- or non-metaphysical thinkers. The clearest metaphysical heritage was seen in Kant's division of the world into nature and freedom and in his attempt to deduce the existence of freedom from the adherence of our will to a noumenal realm that transcends the physical world.

But the larger question hovering in the background has been to what extent the replacement of this view of the world is itself, if not metaphysical, at least insufficiently immanent. I thus suggested that both Korsgaard's view and (the revised) Rawlsian view are insufficiently attentive to the world of injustice and the concrete suffering that goes on here and now because they let their respective theories rest at the level of the mere possibility of justice which is still removed from a theory which seeks to account for the perspective of the loser in the ongoing struggle for justice. The possibility of justice, in other words, treats all actors equally— those who think society adequately provides for them as well as those who do not think this is the case. Without taking account of this relation of

inequality among those addressed by the theory, the theory misses its target, namely those who would do something to improve their own situation and the situation of others.

A speculative theory, as I have already suggested, would seek to tell the story of the conflicting demands that society places on its members by understanding the relation of individual and society as a dialectic between the possibility of freedom and justice and the barriers to the realization of said freedom and justice. A speculative theory gains its anchoring from the possibility of recognition which it shares with the views of Rawls and Korsgaard, but seeks to augment or give a more detailed description of the concrete norms which are needed in order to satisfy the desires of the different members of society. The concrete norms generated by the dialectic of individual and society are historically contingent because they are the products of particular actions of particular agents. As such, their universality consists only in the fact that they have been endorsed by certain members of society as an expression of their freedom. But this means that they are also contingent manifestations of the possibility of freedom that has not yet been realized. A speculative theory would aim not only to retell the concrete decisions made by individuals but would also illuminate the dialectical process of reason giving-itself by showing that justice is always both a contingent social arrangement subject to revision and a genuine expression of the society's ability to be free.

BIBLIOGRAPHY

- Adorno, Theodor. *Minima Moralia. Gesammelte Schriften*. Frankfurt: Suhrkamp, 1997.
- . *Drei Studien zu Hegel. Gesammelte Schriften*. Frankfurt: Suhrkamp, 1997.
- Allison, Henry. *Kant's Theory of Freedom*. Cambridge: CUP, 1990.
- Ameriks, Karl. "Kant's Deduction of Morality and Freedom". *Interpreting Kant's Critiques*. Oxford: OUP, 2003.
- Baynes, Kenneth. *The Normative Grounds of Social Criticism*. Albany: State University of New York Press, 1992.
- Blanchowicz, James. "Reciprocal Justification in Science and Moral Theory". *Synthesis*. 10 (1997) : 447-68.
- Brandt, Richard. "The Science of Man and Wide Reflective Equilibrium". *Ethics*. 100 (Jan. 1990) : 71-90.
- Calvino, Italo. "The Black Sheep" in *Numbers in the Dark*. trans. Tim Park. NY: Vintage, 1996.
- Daniels, Norman. "Wide Reflective Equilibrium and Theory Acceptance in Ethics". *The Journal of Philosophy*. 76 (1979) : 256-82.
- Dworkin, Ronald. "The Original Position". *Reading Rawls*. ed. Norman Daniels. Oxford: Blackwell, 1975.
- G. A. Cohen. "Reason, humanity, and the moral law". *The Sources of Normativity*. Cambridge: CUP, 1996.
- Gauthier, David. "The Unity of Reason; A Subversive Reinterpretation of Kant". *Moral Dealings*. Ithaca, NY: Cornell, 1990.
- Habermas, Jürgen. "Diskursethik— Notizen zu einem Begründungsprogramm". *Erläuterung zur Diskursethik*. Frankfurt: Suhrkamp, 1991.
- Hardimon, Michael. *Hegel's Practical Philosophy*. Cambridge: Cambridge UP, 1994.
- Hare, R. M. "Rawls's *A Theory of Justice*". *Reading Rawls*. ed. Norman Daniels. Oxford: Blackwell, 1975.

- Harris, H. S.. "The Problem of Kant". *Bulletin of the Hegel Society of Great Britain*. 19. (1989): 18-27.
- Hartmann, Klaus. "Hegel; A Non-Metaphysical View". *Hegel; A Collection of Critical Essays*. ed. Alasdair MacIntyre. Garden City, NY: Anchor Books, 1972.
- Hegel, G. W. F. *Elements of the Philosophy of Right*. trans. H. B. Nisbet. ed. Allen Wood. Cambridge: CUP, 1991.
- . *Phenomenology of Spirit*. Trans. A. V. Miller. Oxford: OUP, 1977.
- . *Grundlinien der Philosophie des Rechts*. ed. Eva Moldenhauer and Karl Markus Michel. Frankfurt: Suhrkamp, 1993.
- . *Phänomenologie des Geistes*. Werke 9. Rheinisch-Westphälische Akademie der Wissenschaft. Hamburg: Meiner, 1980.
- . "Über die wissenschaftlichen Behandlungsarten des Naturrechts". Werke 4. Rheinisch-Westphälische Akademie der Wissenschaft. Hamburg: Meiner, 1968.
- Heidegger, Martin. *Sein und Zeit*. Tübingen: Niemeyer, 1993.
- Henrich, Dieter. "Die Deduktion des Sittengesetzes". *Denken im Schatten des Nihilismus*. ed. Alexander Schwan. Darmstadt: Wissenschaftliche Buchgesellschaft, 1975.
- . "Ethik der Autonomie". *Selbstverhältnisse*. Stuttgart: Reclam, 1982.
- Herman, Barbara "The Practice of Moral Judgment". *The Practice of Moral Judgment*. Cambridge: Harvard UP, 1993. pp. 73-93.
- Höffe, Otfried. *Immanuel Kant*. Munich: C. H. Beck, 2000.
- . „Kants kategorischer Imperativ als Kriterium der Sittlichkeit“. *Ethik und Politik*. Frankfurt: Suhrkamp, 1979.
- Hume, David. *Treatise on Human Nature*. ed. David Fate Norton and Mary J. Norton. Oxford: OUP, 2001.
- Itling, Karl-Heinz. "Der naturalistische Fehlschluß bei Kant". *Zur Rehabilitierung der praktischen Philosophie*. ed. Manfred Riedel. Freiburg: Rombach, 1972.
- Kant, Immanuel. *Grundlegungen der Metaphysik der Sitten*. Hamburg: Meiner, 1999.
- . *Practical Philosophy*. ed. and trans. Mary Gregor. Cambridge: CUP, 1996.

- . *Kritik der reinen Vernunft*. Hamburg: Meiner, 1998.
- Kaulbach, Friedrich. *Das Prinzip Handlung in der Philosophie Kants*. Berlin: de Gruyter, 1978.
- Korsgaard Christine. *Locke Lectures*.
<http://www.people.fas.harvard.edu/~korsgaar/#Locke%20Lectures>
- . "Kant and Aristotle on Morally Good Action". *Aristotle, Kant, and the Stoics*. ed. Stephen Engstrom and Jennifer Whiting. Cambridge: CUP, 1996.
- . "Kant's Analysis of Obligation". *Creating the Kingdom of Ends*. Cambridge: CUP, 1996.
- . "Kant's Formula of Humanity". *Creating the Kingdom of Ends*. Cambridge: CUP, 1996.
- . "Morality as Freedom". *Creating the Kingdom of Ends*. Cambridge: CUP, 1996.
- . "Reasons We Can Share". *Creating the Kingdom of Ends*. Cambridge: CUP, 1996.
- . "Self-Constitution in the Ethics of Plato and Kant". *The Journal of Ethics*. 3 (1999) : 1-29.
- . "Skepticism about Practical Reason". *Creating the Kingdom of Ends*. Cambridge: CUP, 1996.
- . "The Myth of Egoism". The Lindley Lectures. University of Kansas, 2000.
- . "The Normativity of Instrumental Reason". *Ethics and Practical Reason*. eds. Garrett Cullity and Berys Gaut. Oxford: Clarendon Press, 1997.
- . *The Sources of Normativity*. Cambridge: CUP, 1996.
- Lenk, Hans. "Freies Handeln als Interpretationskonstrukt". *Kant in der Diskussion der Moderne*. ed. Gerhard Schönrich and Yasushi Kato. Frankfurt: Suhrkamp, 1996.
- McDowell, John. "The Role of *Eudaimonia* in Aristotle's Ethics". *Mind, Value and Reality*. Cambridge, MA: Harvard, 1998.
- Nagel, Thomas. "Moral Luck". *Mortal Questions*. Cambridge: CUP, 1979.
- . "Rawls on Justice". *Reading Rawls*. ed. Norman Daniels. Oxford: Blackwell, 1975.

- Nussbaum, Martha. "Bernard Williams: Tragedy and Justice". Conference Paper Georgia State University, March 17-18, 2006.
- . *The Fragility of Goodness*. Cambridge: CUP, 2001.
- O'Neill, Onora. "Construction in Rawls and Kant". *The Cambridge Companion to Rawls*. ed. Samuel Freeman. Cambridge: CUP, 2003.
- . "Ethical Reasoning and Ideological Pluralism". *Ethics*. 98 (1988) : 705-722.
- . "The Method of *A Theory of Justice*". *John Rawls: Eine Theorie der Gerechtigkeit*. ed. Otfried Höffe. Berlin: Akademie Verlag, 1998.
- Parfit, Derek. *Reasons and Persons*. Oxford: OUP, 1986.
- Patten, Allen. *Hegel's Idea of Freedom*. Oxford: OUP, 1999.
- Pieper, Annemarie. "Wie ist ein kategorischer Imperative möglich?". *Grundlegung zur Metaphysik der Sitten; Ein kooperativer Kommentar*. ed. Otfried Höffe. Frankfurt: Klostermann, 2000.
- Pinkard, Terry. "How Kantian Was Hegel?". *Review of Metaphysics*. 43 (1990) : 831-38.
- . "Innen, Außen und Lebensformen: Hegel und Wittgenstein". in *Hegels Erbe*. ed. Christoph Halbing, Michael Quante and Ludwig Siep. Frankfurt: Suhrkamp, 2004 : 254-94.
- . "Self-understanding and Self-Realizing Spirit in Hegel's Ethical Theory". *Philosophical Topics*. 19.2 (1991) : 71-98.
- . "The Categorical Satisfaction of Self-Reflexive Reason". *Bulletin of the Hegel Society of Great Britain*. Vol. 19 (1989) : 5-17.
- . "The Successor to Metaphysics". *The Monist*. 74.3 (1991) : 295-328.
- . *Hegel's Phenomenology*. Cambridge: 1994.
- Pippin, Robert. "Avoiding German Idealism: Kant, Hegel, and the Reflective Judgment Problem". In *Idealism as Modernism*.
- . "Hegel and Institutional Rationality". *The Southern Journal of Philosophy*. 39 (2001) 1-25.
- . "Hegel on the Rationality and Priority of Ethical Life". *Neue Hefte für Philosophie*. 35 (1995): 95-126.

- . "You Can't Get There from Here: Transition Problems in Hegel's *Phenomenology of Spirit*". in *The Cambridge Companion to Hegel*. ed. Frederick Beiser. Cambridge: CUP, 1993.
- . "Brandom's Hegel". *European Journal of Philosophy*. 13:3 (2005) : 381-408.
- . *Hegel's Idealism*. Cambridge: CUP, 1989.
- Prauss, Gerold. *Kant über Freiheit und Autonomie*. Frankfurt: Klostermann, 1983.
- Quante, Michael. *Hegels Begriff der Handlung*. Stuttgart: Frommann-Holzboog, 1993.
- Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard-Belknap, 1999.
- . "Kantian Constructivism in Moral Theory". *Collected Papers*. ed. Samuel Freeman. Cambridge, MA: Harvard, 2001.
- . *Lectures on the History of Moral Philosophy*. ed. Barbara Herman. Cambridge, MA: Harvard, 2000.
- . "Themes in Kant's Moral Philosophy". *Collected Papers*. ed. Samuel Freeman. Cambridge, MA: Harvard, 2001.
- . *Political Liberalism*. New York: Columbia, 1996.
- Rorty, Richard. "Comments on Robert Pippin's 'Naturalness and Mindedness: Hegel's Compatibilism'". *European Journal of Philosophy*. 7:2. (1999) : 213-16.
- Rüdiger Bittner. „Maximen“. *Akten des 4. Internationalen Kant-Kongresses. Mainz 6-10 April 1974*. Teil II.2 Hg. Gerhard Funke. pp. 485-98.
- Sandel, Michael. *Liberalism and the Limits of Justice*. Cambridge: CUP, 1998.
- Scanlon, Thomas. "Rawls on Justification". *The Cambridge Companion to Rawls*. ed. Samuel Freeman. Cambridge: CUP, 2003.
- Scanlon, Thomas. *What We Owe Each Other*. Cambridge, MA: Harvard-Belknap, 1998.
- Schönecker, Dieter and Allen Wood. *Kants 'Grundlegung zur Metaphysik der Sitten'; Ein einführender Kommentar*. Paderborn: Schöningh, 2002.
- Susan Moller Okin. "Political Liberalism, Justice, and Gender". *Ethics*. 105.1 (1994) : 23-43.
- Taylor, Charles. *Hegel*. Cambridge: CUP, 1975.

Watkins, Eric and William Fitzpatrick. "O'Neill and Korsgaard on the Construction of Normativity". *Journal of Value Inquiry*. 36.2-3 (2002) : 349-67.

Williams, Bernard. "Internal and External Reasons". *Moral Luck*. Cambridge: CUP, 1980.

---. "Moral Luck". *Moral Luck*. Cambridge: CUP, 1980.

---. "Persons, Character and Morality". *Moral Luck*. Cambridge: CUP, 1980.

---. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard, 1985.

Wood, Allen. *Hegel's Ethical Thought*. Cambridge: CUP, 1990.

---. *Kant's Ethical Theory*. Cambridge: CUP, 1999.