

DUTIES OF A FREE PERSON

by

Brian Arsenault

Submitted in partial fulfilment of the requirements
for the degree of Master of Arts

at

Dalhousie University
Halifax, Nova Scotia
August 2012

© Copyright by Brian Arsenault, 2012

DALHOUSIE UNIVERSITY

DEPARTMENT OF PHILOSOPHY

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled “DUTIES OF A FREE PERSON” by Brian Arsenault in partial fulfilment of the requirements for the degree of Master of Arts.

Dated: August 24, 2012

Supervisor: _____

Readers: _____

DALHOUSIE UNIVERSITY

DATE: August 24, 2012

AUTHOR: Brian Arsenault

TITLE: Duties of a Free Person

DEPARTMENT OR SCHOOL: Department of Philosophy

DEGREE: MA CONVOCATION: October YEAR: 2012

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions. I understand that my thesis will be electronically available to the public.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

The author attests that permission has been obtained for the use of any copyrighted material appearing in the thesis (other than the brief excerpts requiring only proper acknowledgement in scholarly writing), and that all such use is clearly acknowledged.

Signature of Author

For those who choose to paddle upstream

TABLE OF CONTENTS

ABSTRACT.....	vi
LIST OF ABBREVIATIONS USED.....	vii
ACKNOWLEDGEMENTS.....	viii
CHAPTER 1 INTRODUCTION.....	1
CHAPTER 2 KORSGAARD’S CONSTITUTIVISM.....	7
2.1.....	7
2.2.....	13
2.3.....	15
2.4.....	22
2.5.....	30
2.6.....	34
CHAPTER 3 FREEDOM, NORMATIVITY AND SKEPTICISM.....	37
3.1.....	37
3.2.....	38
3.3.....	41
3.4.....	47
3.5.....	52
3.6.....	56
CHAPTER 4 THE EMPIRICIST ALTERNATIVE.....	65
4.1.....	65
4.2.....	70
4.3.....	74
4.4.....	79
CHAPTER 5 CONCLUSION.....	81
5.1.....	81
5.2.....	83
5.3.....	86
BIBLIOGRAPHY.....	92

ABSTRACT

The following is an attempt to ground personal duty – duty which is both believed and felt by all agents. To do this, I look at two contrasting attempts. The first is a rationalist attempt, which tries to ground it in conceptual necessity. The particular rationalist theory I critique is that of Christine Korsgaard, which tries to extract normativity from the constitutive requirements of action: it is the view that, in virtue only of the fact they are acting, agents are drawn toward one type of action in particular. To this, I argue that "action" and "agency" are too thin as concepts to generate the kind of normativity that Korsgaard is trying to get from them. In general, the rationalist, by appealing to conceptual necessity, only illuminates how this or that thing is done correctly, and not that this or that thing is what ought to be done. The second type of attempt to ground duty is an empiricist one, which uses empirical fact rather than conceptual necessity as its basis. In particular, it uses contingent facts about the things which *are* agents (people, for example), and what makes them feel a sense of duty. I argue that, ultimately, it is this type of grounding of duty which can be successful.

Throughout, I emphasize two crucial points. The first is the freedom of the individual, in particular to pursue her own ends by her own means, and for whichever reasons she chooses. Therefore, the eponymous free person is represented by a harsh skeptic, who is reasonable, in the sense of being open to rational persuasion, but highly resistant to having duty imposed on him. Mainly, he asserts his freedom to pursue his own goals, and even go against his own wants, if he feels he should. Duties of a free person are thus those that the skeptic would freely pursue, perhaps even against his own self-interest. This is the second crucial point: that duty is not a "want" or "desire;" rather, it is quite often what one does *against* one's own wants or desires. This is an empirical truth of what occurs when one acts according to duty, and is something which no non-trivial definition of desire can adequately capture.

It is with these two things in mind that the empiricist project must convince the skeptic of his duty. I argue that a paradigmatic example of the project doing so lies in Harry Frankfurt's argument that we cannot help but act in the interests of that which we love. Love is a case in which the skeptic would put others' interests above his own, and do so freely and without regret. But while I take Frankfurt's as a paradigm case of convincing the skeptic, I leave the door open for other possibilities. Then, I close with some brief remarks about the connection between moral duty and the duty of the skeptic, and about some implications of my arguments with respect to conation and cognition.

LIST OF ABBREVIATIONS USED

FPL	Dalton, D. (1994) "Freedom: The Philosophy of Liberation"
EPF	Dick, D. (2009) <u>Ethics and the Possibility of Failure: Getting it Right About Getting it Wrong</u>
AS	Enoch, D. (2006) "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action"
ANL	Frankfurt, H. (1999) "Autonomy, Necessity and Love"
FGS	Frankfurt, H. (1999) "On Caring"
NIR	Korsgaard, C. (1997) "The Normativity of Instrumental Reason"
SC	Korsgaard, C. (2009) <u>Self-Constitution: Agency, Identity and Integrity</u>
NMAT	Rosen, G. (Manuscript) "Normativity, Meaning and All That"
AAPW	C. Korsgaard's argument against particularistic willing
CI	The Categorical Imperative

ACKNOWLEDGEMENTS

I would like to thank all those who have allowed me to reach this point, which is one of both completion and transition. Academically, I could not be here were it not for the encouragement and support of David Dick, Ishtiyaque Haji, John Baker, David Hitchcock, James Hammond, and Kathleen Garay. As for the present enterprise, I would like to thank Duncan MacIntosh for agreeing to supervise the preparation of this work, and for his continued and invaluable feedback throughout the process of writing it. And also, I must thank Greg Scherkoske and Chike Jeffers for agreeing to subject themselves to it.

This past year has been both fulfilling and enjoyable thanks largely to my fellow MA candidates Kevin Phang, Katie Stockdale and Jasmine Smart, in all of whom I have made lifelong friends. I would also like to acknowledge in this regard the graduate environment in the Department of Philosophy at Dalhousie: profs, alumni and PhD candidates past and present. I would also like to thank Dalhousie University and its Department of Philosophy for accepting me into the MA program and allowing me this opportunity.

I would also like to thank Bob Stewart for helping to turn my analytical mind in a philosophical direction, for his sustained and continued friendship over the years, and for teaching me that there are no sacred cows. Without our beer-driven, frost-bitten, philosophical porch chats, I truly would not be the man I am today, and I'd probably have a lot more money.

While I am a believer of fierce independence at a personal level, no one does it all alone. I would like to thank all those who have helped me and shown me patience and guidance – parents, friends, partners, teachers, employers, colleagues. There have been many such positive influences and there will likely be many more; I am truly grateful and fortunate to be affected by them.

CHAPTER 1 INTRODUCTION

There is a strong tension between one being free to do as one likes and one having a duty one ought to fulfill. One component of being a free person, we might say, is being free from obligation, free from one's duties toward others and toward society, free to do as one pleases. Duties, we might think pre-theoretically, act as *constraints* on one's freedom, in that they prevent or supersede the fulfillment of one's own chosen wishes. This is especially true when these duties are imposed on one from some source external to oneself with which one does not agree, such as a societal norm; in such cases, acquiescence is done, if at all, begrudgingly, perhaps in fear of the consequences of noncompliance. (Mill's phrase, "tyranny of the majority," comes to mind here.) But a maximally free person is, at the very least, free from fear and acquiescence, free from constraints on her liberty. And being thusly free, it might be difficult to establish that she in fact has any duty, since that would seemingly involve the prioritization of something else over her own goals and preferences.

But despite all of this, in this thesis my aim is to show how a truly free person can feel a sense of duty and thus act as she should rather than how she likes. That is, I aim to show that a person can feel neither constrained nor coerced and still feel obliged toward something outside of herself. This is very much in line with a tradition in philosophy that emphasizes control (particularly *rational* control) over oneself for the sake of some "greater good." However, as my opening remarks suggest, a free person need not accept the "greater" good as her own good, lest she feel constrained from something outside of herself. So the project of establishing duties of a free person is one of arguing for duties

to which she herself *freely assents*, with the control she exerts being for the sake of a duty which is freely chosen. What kind (or kinds) of duty would such a free person choose?

To make the case for such duties as forcefully as possible, I will be using an extremely skeptical position as the standard for freedom. One of the reasons for this is that the skeptic, as I will define him, pronounces (among other things) the same skepticism toward duty that David Hume famously did when he refused to accept the validity of any argument with normative conclusions and non-normative premises (also known as "deriving an ought from an is"). But like Hume, I present the skeptic as being reasonable, if a bit fastidious: he is hard to convince, but not impossible; he is open to argument and evidence, but will grant no charity in doing so. And most of all, he asserts and defends his freedom. The skeptic is therefore meant to represent freedom's last stand against duty: if he can be convinced then anyone can. The duties of a free person are those which are had by anyone.

My strategy will be to examine two kinds of attempts to ground duty, attempts which lie in "rationalist" and "empiricist" theories. A rationalist tries to ground duty by appealing to a kind of conceptual necessity: to act means to do this or that, and in virtue of the fact that this is the activity in which one is engaged, and in virtue of how that activity is done *well*, one feels the normative pull in a certain direction. A skeptic is therefore supposed to recognize this necessity and be gripped by it. He cannot reasonably deny that he is *acting*, after all, which means (on this view) being guided by some principles of action. The rationalist thus convinces the skeptic that by going against his duty he is going against what he himself has set out to do by acting, something a reasonable person would not want to do. And since I do mean to present the skeptic as

being reasonable, going against himself in this way would be unacceptable. So if the rationalist can convince him that going against his duty is in fact going against himself, then she would have successfully convinced him of his duty. This is the strategy behind the theory of Christine Korsgaard, which is the subject of the chapter 2.

Contrastingly, an empiricist theory tries to ground duty by appealing to contingent, empirical facts about the self. In its classical form, most prototypically in Hume, duty consists in taking the means to one's own chosen ends.¹ The result of acting contrary to duty is supposed to be the same as it was in the rationalist picture: it amounts to going against oneself and what one is trying to do. In the empiricist picture, however, it is not some conceptual necessity that provides the normative pull, but one's own desires. The skeptic is therefore motivated by these desires, and his rationality and beliefs about the world guide him in fulfilling them. But as I will argue, one's "desires," in any real sense, do not constitute the only basis by which one makes a decision. And in fact, acting according to one's duty can quite often mean going *against* one's desires in favour of what one deems right. In other words, I will be arguing that the claim that one's desires pull one toward one's duty is empirically false – that duty, instead, involves the *superseding* of one's desires.

The next chapter, as I said, is on Korsgaard's rationalist picture. Hers is a "constitutivist" theory, meaning that she tries to convince the skeptic by appealing to what is constitutive of action, or what it essentially involves. In chapter 3, I will use the lessons learned from the analysis of Korsgaard to outline a framework for freedom and

¹ This, for Hume, is what is *rationally* required. *Moral* duty involves the "common point of view" which is the result of certain empirical facts such as the "moral sentiments" that we share. In chapter 4, I will be arguing for a position in which the skeptic is gripped by a duty grounded in empirical fact, but it is different from this view of Hume.

duty and how they go together. This includes giving a precise formulation of the skeptical position and the freedom it demands, as well as giving an account of the kinds of normative inferences to which the skeptic can agree. (If not normative conclusions from non-normative premises, then what?) Also, with this framework, we will be in a better position to say more precisely just why the rationalist project failed to ground the duty of the skeptic.

After the framework for freedom and duty is outlined, chapter 4 will approach the issue from an empiricist perspective, while keeping in mind the empiricist's problem with duty that I just mentioned. The key question I aim to answer is that of whether some empirical fact about a person's make-up can be such that a free person will agree to some duty based on it. I will present a theory that I believe characterizes just this kind of fact, while avoiding Hume's problem of deriving an "ought" from an "is." But importantly, I do not mean the example I cite to exhaust all possibilities for such facts. Rather, I take myself as having outlined precisely how the skeptic can assent to a duty, and then as having presented *one* kind of duty to which he assents, leaving the door open for other possibilities.

Throughout, I will be emphasizing the importance of the connection between theoretical entities and the things which they purport to characterize. This is important to place duty within the framework of real human experience. For example, as I said, the argument that I will be putting forth against the classical empiricist picture was based on an empirical claim, that duty is not necessarily (and is quite often not) done in alignment with one's desires, where this term means something like one's "wants". But there is a sense in which one might say that it is analytically true that *all* action is done in

alignment with one's desires, since only desires can move one to action. But if this is true, then what does the term “desire” denote? It seems that the things which *actually* move us to action are quite disparate in kind, too much so to be captured only by one term. I will argue that in any non-analytic sense, the thesis that only desires move one to action is empirically false. The rationalists contend that one can do one's duty *in spite* of one's desires, which seems (perhaps ironically) truer to the empirical reality of decision-making. However, as I said, the rationalist relies on a conception of action (as well as, we shall see, of agency) which brings normativity with it, and her difficulty will be in giving such a characterization that maintains its connection with the things which actually are actions – that is, with the things it purports to describe. But in emphasizing the importance of a theory accurately representing the things in the world of which it purports to treat, I do not mean to stack the deck in favour of one kind of theory over another. Rather, the world is ultimately a theory's (or at least certainly a theory of *action's*) domain of description – or, if it is a normative theory, its domain of *prescription* – and so its descriptions and prescriptions must map onto the world somehow. And further, the skeptic will only be convinced of his duty if the entity which a theory demonstrates as having a duty in fact corresponds to him somehow.

Establishing duties of a free person I see as an uphill struggle. As I said, I mean to present the skeptic forcefully in order to bring as many doubters into the fold as possible. Doubting, after all, is a reasonable thing to do. It is the basis for much of philosophy, the paradigmatic (I hope) discipline of reasonable people. It is important when speaking of duty that any reasonable person can be made to agree, since otherwise she is determined by forces outside of herself. Nonetheless, there are duties that all

reasonable people can agree on, whether or not these come from reason alone, and they form the basis for our common lived experience. My goal is to give an account of such duties.

CHAPTER 2 KORSGAARD'S CONSTITUTIVISM

In this chapter, I wish to analyze a rationalist (Kantian) conception of normativity which is to be contrasted, later, with an empiricist conception. The reason for this is to present a case study of the *ways* in which certain problems in the philosophy of action are addressed, and the *kinds* of problems the theories that purport to address them run into. In this way, I wish to hit the ground running. By seeing what a theory is trying to do and how it is trying to do it, we can get a good idea of the issues involved. The rationalist theory I wish to scrutinize in this first chapter is Christine Korsgaard's constitutive theory of agency, outlined most fully in her Self-Constitution (2009).² I will argue that her theory is ultimately unsuccessful, but there are many valuable lessons to be learned from it – both from its virtues and its vices. In particular, one of its virtues is the *way* in which it argues for its normative standard – that is to say, the way it tries to convince those who would be skeptical of it – and I will be making use of this method in later chapters when developing my own account of duty.

2.1

Speaking very generally, a theory will have certain problems it is trying to address and a theoretical apparatus that it introduces or employs. The theory may or may not solve these problems, and it may or may not give rise to new ones. I take Korsgaard's theory to be an attempt to solve four problems in particular. Firstly, it addresses what David Dick

² Korsgaard, C. (2009) This is hereafter referred to in the text as SC with page numbers.

(2009),³ in his treatment of Korsgaard, calls the “mereological problem”. As Korsgaard says,

to regard some movement of my mind or my body as *my action*, I must see it as an expression of my self as a whole, rather than a product of some force that is at work *on* me or *in* me. Movements that result from forces working *on* me or *in* me constitute things that happen to me. To call a movement a twitch, or a slip, is at once to deny that it is an action and to assign it to some part of you that is less than the whole: the twitch to your eyebrow, and the slip, more problematically, to your tongue. (SC 18)

The mereological problem is thus how an agent can come to be, and come to perform an action, in a unified way. The alternative to this is either that one identify the agent with some smaller subset of a person – say, with her beliefs and desires – or that one includes twitches, and the like, in the category of genuine actions. Korsgaard finds this unacceptable, arguing that action requires a whole agent “over and above” the “mere heaps” of desires, incentives and so on, and thus aims to solve the mereological problem.

The second problem Korsgaard aims to solve is that of bad actions. In “The Normativity of Instrumental Reason,”⁴ one of Korsgaard's arguments against the empiricist conception of practical rationality is as follows: “Hume identifies a person’s end as what he wants most, and the criterion of what the person wants most appears to be what he actually does. The person’s ends are taken to be revealed in his conduct.” (NIR 230) If this is true, and if rationality consists in taking the means to one's ends, then on

³ Dick, D. (2009) Ethics and the Possibility of Failure: Getting it Right About Getting it Wrong. Hereafter referred in the text as EPF with page numbers.

⁴ Korsgaard, C. (1997) This is hereafter referred to in the text as NIR with page numbers.

Korsgaard's understanding of Hume there is no bad action, with “bad” in this case meaning “irrational”. Persons, then, are *always* rational on this account, and so the account must be problematic. Korsgaard's account in Self-Constitution is an attempt to characterize the structure of rational action in such a way that agents can fail to abide by it. Clearly, sometimes we screw up.⁵

Thirdly, and perhaps most centrally for Korsgaard, she is attempting to account for the normativity of action, both moral and rational. Normativity has two separate functions, which are to guide and to motivate; in other words, a norm is a standard to which we both aspire and conform. As Korsgaard puts it, “the *normativity* of a law or demand [is] the grounds of its authority and the psychological mechanisms of its enforcement, the way that it binds you.” (SC 2) In NIR, Korsgaard argues extensively that both the “empiricist” and “rationalist” conceptions of action, as she characterized them, failed to give an account of both of these simultaneously. Her characterization of the empiricist picture, as I said, entails that it fails to allow for bad actions, and therefore

⁵ Korsgaard's reasons for criticizing the Humean account on this ground can of course be contested, in particular by one invoking the distinction between revealed and expressed preferences. As I will be arguing throughout this thesis, what is important are the real factors that move one to action – a person's wants and urges, for example – as well as the control one exerts over them and the reasons one does so. Korsgaard does well to capture the lurking triviality in the Humean account: if a preference is simply what moves one to action, then it is analytically true that one's actions align with one's preferences. The expressed/revealed preference distinction tries to address this triviality. I will be arguing later against the use of the concept “desire” since I see it as lacking any concrete denotation, and I see “preference” as having the same issue. The expressed/revealed preference distinction I see as an *epistemic* methodology for determining one's “genuine preferences.” That is, we have the evidence of the agent's behavior, and we have the evidence of what she *says* she prefers, so what, given *all* the evidence, does she “really” prefer? But these genuine preferences, again, seem to have no concrete denotation and succumb to the charge of being trivial; they describe the choice that the causal factors within the agent yield rather than any of the factors themselves. Thus, “I chose this because I preferred it” is trivial the same way that saying opium puts one to sleep because of its dormative property is trivial. “Preference” (genuine or otherwise) might be a useful theoretical concept – in economics, for example (since that discipline is concerned with choices and not with a rich description of the agents who make them) – but it does not correspond to any “facts on the ground.” While this is how I see the expressed/revealed preference discussion going here, this argument would probably need to be stated a bit more fully to gain wide acceptance.

fails to act as a guide. Her characterization of the “dogmatic rationalist” picture, by contrast, is that it “show[s] how the instrumental principle can guide us. But it does not show why we must be motivated to follow that guide.” (NIR 242) The normativity she derives in her theory is expressly intended to fulfill both of these roles: she wishes to provide the norms of action that both guide and motivate, but, importantly, in such a way that we can fail to abide by them.

Fourthly, related to the third point, Korsgaard wants her theory to stand up to skeptical challenges. I will outline three such challenges below (two in this chapter, and one more in the next), but it is worth describing one of them briefly so that I may properly characterize the problem Korsgaard is addressing. A skeptic might buy into a certain theory of rational action, meaning that he accepts the criteria it outlines for what makes an action rational, but still he might ask, “why should I do this? Sure, it is the rational thing to do, but so what?” Indeed, this may have some resonance with everyday life: we know what the rational thing to do is, but we could not be bothered, or we would rather not, or maybe we are still not convinced that we *ought* to do it. Korsgaard recognizes this and aims her theory at addressing it. Her solution, in broad strokes, is to have normativity, with both of its functions, result directly from what is *constitutive* of action, and so for a skeptic to ask that question is for him to ask why he should act at all; the standard for action comes from what it *means* to act. “The idea of a constitutive standard is an important one, for constitutive standards meet skeptical challenges to their authority with ease.” (SC 29)

So this is what I take Korsgaard to be addressing with her theory. Now the task is to see how she goes about doing this, whether she solves the problems she is attempting to solve, and whether any new ones pop up.

One type of “new” problem I will be keeping my eye out for is whether Korsgaard allows for an acceptable conception of autonomy. It is very fair to expect her theory to do this since she explicitly states, in keeping with her Kantian roots, that one must see oneself as the cause of one's actions; indeed, this is largely the reason that she feels the need to solve the mereological problem. The word “autonomy”, when broken into its component parts, entails that one is governed by a law of one's own, or that one is “self-determined”. I think Korsgaard would appreciate having her feet held to the fire with regard to accommodating this indispensable feature of human action, and I intend to do so. In case it was not clear above, the skeptic, to which I take Korsgaard to be responding, very much brings this issue to the fore, asserting his freedom to go against whichever norms or dictates a theory puts forward. Therefore, in response, the theory must satisfy the skeptic with an acceptable amount of freedom while at the same time convincing him to conform or obey.

Another type of problem I will watch out for is the following. A philosophical theory about anything must hold up to two forms of scrutiny. Firstly, and most obviously, it must be internally consistent (“paraconsistency” notwithstanding); this is really a bare minimum requirement. But secondly, the theory must tether itself to reality somehow if it is actually to be *about* anything. For example, a Humean “desire” is certainly *somehow* related to what one would colloquially call a “want”, but it is clearly more expansive as well; for example, Hume’s “calm passions” (Hume, 1739, II.iii.3)

fulfill the theoretical function of a desire without being very much like something one would call a “want.” What, then, *is* a desire, exactly? And further, what would a theory prescribing (for instance) that one advance one’s desires really be prescribing? For example, one might think to oneself, “I really *want* to do this, but I don’t feel that it is right. And I’m trying to be rational, which means [according to the theory] advancing my desires. So, what ought I rationally to do?” Is the *want* a desire, or the feeling that the action is not right, or both, or neither? The answer to this question will obviously weigh heavily on what the theory in question prescribes.⁶ It is a problem for a theory to have this property, where so much hangs on an imprecisely defined concept, because it makes it unclear just what the theory describes or prescribes (depending on what kind of theory it is).

To avoid this, a theory should be rooted in empirical reality. If “desire” is not related closely enough to something that is actually had by a human being, then what good is a theory which employs it, no matter how consistent, elegant or parsimonious it may be? The same goes for “agent” and its relationship to “person”, and several other theoretical entities and their real-world analogs. One frequent strategy for argumentation in the philosophy of action is to argue that such-and-such is inconsistent with our experience as it actually occurs; for example, an argument might say that it cannot be that only desires motivate, since we quite often *fail* to act on our wants (e.g., due to fear), and so something else has motivated us otherwise. But this is an acceptable form of argumentation only if the entities in the theory (desires, in this example) purport to be either identical, or at least strongly related, to entities in the real world and say something

⁶ This is similar to the charge that the verdict of a Kantian maxim is determined by its formulation, that how an action is described determines whether it follows the categorical imperative.

substantive about it and how we experience it. I hope that this is an obvious point, but it is one worth keeping in mind, and it is one to which I will return.

2.2

I will proceed with the analysis of Korsgaard's theory in a somewhat backwards fashion, introducing my criticism of her view before expounding it fully. The reason for this is that two separate (although similar) interpretations of what is going on in Korsgaard's book are possible, each of which addresses problems raised by the other, but which are not compatible. If we accept one and not the other, then the other's problems remain, and vice versa. Thus, I will frame her overall theory in the form of a dilemma, with each of these interpretations as the dilemmatic disjuncts, and derive the undesirable consequences of each one. Then, I will see if an attempt can be made to avoid the dilemma and get all of the positive results that Korsgaard is looking for.

Korsgaard's view, in slogan form, is that action is self-constitution. One “becomes oneself” through one's actions, and the constitutive aims of action provide normativity (as I described it above) toward a certain class or type of action in particular. Essentially, the dilemma I am using to frame my characterization of her theory boils down to the following disjunction: either bad actions constitute an agent or they do not. If they do, then performers of bad actions, being thusly constituted, have no further impetus to perform good actions. Bad actions, being actions, meet the constitutive aims of action, and in so doing constitute the agent, who has thus met her aim of self-constitution. Therefore normativity toward *good* action, either as a guide or an incentive,

is extraneous. Or put another way, the skeptic will happily constitute himself with bad actions and assert that nothing is gained further by performing good ones.

On the other hand, if bad actions do *not* constitute agents, then they do not meet the constitutive criteria for action, and therefore they are not actions at all. And further, since action is self-constitution, the things that perform them are not even agents. On this interpretation of Korsgaard, when someone performs what we would ordinarily call a bad action, we must characterize it in one of two ways. Firstly, we might say that it is not really “bad”. Since Korsgaard is shooting for both rational *and* moral normativity, this would mean that the seemingly “bad” action is neither immoral nor irrational. Korsgaard would, very rightly, resist this interpretation since it would be easy to develop clear counterexamples. Secondly, we might say that the “bad” action is simply not an action at all. Inasmuch as one is trying to perform an action, one must try to abide by the constitutive standards of action; a failure to conform to those standards is simply a failure to perform an action. Thus, on this interpretation, constitutive standards are normative, in the sense that they are guiding and (we shall grant for the moment) motivating, but failure to live up to them results in “non-actions” and “non-agents”. The only *actions* which are done *by* the agent are good ones, i.e., ones that conform to the standard. Those things which are bad “actions” are done by something else, perhaps things like desires and impulses; agents can never be full participants in something “bad”. (Recall that the refusal to admit of irrational action was one basis for Korsgaard's critique of the empiricist conception of rationality, which should initially steer us away from this interpretation. However, the way in which she outlines her metaphysics of normativity steers us right back toward it.)

So these are the competing interpretations of Korsgaard's theory which form two horns of a dilemma, as I will describe it. Korsgaard's attempt at a resolution will be to say, roughly, that bad actions constitute, but good actions constitute *better*. But this, as I will argue, is not a satisfactory response; either there is an agent there or there is not. But before I get to the point of her response, it is now time to outline her theory, keeping in mind these two interpretations, what I take to be her four main goals, and the two additional aspects (autonomy, connection with reality) with which I am concerned. This is perhaps a lot of balls to juggle at once, but doing so will serve the larger purpose of examining the structure of theories of action in general, of which constitutive theories comprise one type, and of this type Korsgaard's theory is one example. Onward, then, to the business of describing it.

2.3

Korsgaard's theory has several pillars which are meant to serve as its foundation. First and foremost of these is the "necessity of action." An action, in Korsgaard's sense, is "an act done for the sake of an end." (quote in SC 18; the arguments for this are in SC 8ff) An act is the brute physical truth of what is going on; a description of an action, by contrast, incorporates the end for which the act was done. To use a couple of Korsgaard's own examples to explain:

making a false promise and committing suicide are what I am calling "acts";
making a false promise in order to get some ready cash, committing suicide in order to avoid misery are what I am calling "actions." (SC 12)

This is just some stipulative distinction-making, so all is well so far. Now we can rewind to the very beginning of the book (after the preface):

Human beings are *condemned* to choice and action.⁷ Maybe you think you can avoid it, by resolutely standing still, refusing to act, refusing to move. But it's no use, for that will be something you have chosen to do, and then you will have acted after all. Choosing not to act makes not acting a kind of action, makes it something that you do. (SC 1)

As an introductory remark (which it is), this has a certain intuitive appeal. But after her description of what is meant by “action”, the remark becomes somewhat more substantial – i.e., *in Korsgaard's stipulated conception of action*, we are condemned to action. (I mention this only for clarification, withholding judgment on it for the time being.)

The necessity of action leads into a second pillar in Korsgaard's theory, which is the argument against “particularistic willing,” (AAPW) which is a term I will explain shortly. (An explanation of the AAPW is found in SC 72ff; Dick also has a very helpful reconstruction of it in EPF 68-69). Importantly, the AAPW is intimately bound to Korsgaard's goal of solving the mereological problem, in that it is used as a premise that *only* “universal willing” – which is the opposite of particularistic willing – can solve it.

It is crucial for Korsgaard that in all cases of action the whole agent is the cause of his action, seen independently of the causes at play on him or in him. Following Kant,

⁷ Although I could not fit it smoothly into the text, it is worth noting that the phrasing of being “condemned to choice and action” is an homage to Sartre’s “condemned to be free,” and there is also a strong resemblance between Korsgaard’s “self-constitution” and Sartre’s “self-determination.” A principal difference between them is that in Sartre we cannot help but choose, and in Korsgaard we cannot help but act. I believe that my conception in the next chapter of “global freedom” is in line with Sartre’s idea of freedom but, I will argue, not with Korsgaard’s. Thus, the freedom to which one is “condemned” is different between those two philosophers.

Korsgaard believes “that every action involves two factors, an incentive and a principle. The incentive is the thing that presents the action to your mind as eligible; the principle is what determines whether it is to be chosen or not.” (SC 22) An action, by definition, takes the form of a Kantian maxim “I will do act-A in order to promote end-E.” (SC 11) Therefore, “when you determine yourself to be the cause of the movements which constitute your action [i.e., solve the mereological problem], you must identify yourself with the principle of choice upon which you act” (SC 75) – i.e., the maxim of your action.⁸ In doing this, one appeals to the principle of “A’ing in order to promote end-E,” and would will *in all similar situations* that A’ing be done to promote end-E. That is, one takes “A’ing in order to promote end-E” to be a *law* governing oneself. This is what it means to will universally – to act according to such a principle, which one takes to be a law, and thereby extend one’s maxim beyond one’s current action – and this is required in order to solve the mereological problem. Recognizing that this is how action works is the first step in the AAPW.

“The second step is to see that particularistic willing makes it impossible for you to distinguish yourself, your principle of choice, from the very incentives on which you act.” (SC 75) To will particularistically means, by definition, not to see one’s maxim as representative of any type (since then it would be a universal), but rather to “embrace the incentive in its full particularity.” (SC 76) “Thus particularistic willing for particularistic reasons can be thought to set no precedents, appeal to no principles, and speak in no

⁸ The reason for “identifying” oneself with the principle, instead of just following it or complying with it or assenting to it, is related to Korsgaard’s notion of “practical identity.” It is this type of identity which is “constituted” through her titular self-constitution, and she says that actions allow us to pick out something as being identifiable as the agent. Presumably, the agent is also using this to identify *herself* as well, since otherwise she would have no practical identity. It may or may not be necessary for Korsgaard for agents to “identify” with their principles, but this is how she describes it.

universal terms.” (EPF 69) For example, if I am hungry I might go to the fridge to get some food – which is the maxim for my action – because that is (one example of) what one *does* when one wants food in my situation. To will particularistically, I (by definition) do not recognize this as what one *does* in my situation to meet that end, but only as what I *am doing now* to meet it. Faced with the same situation tomorrow, I cannot appeal to this decision to help me with that one since this one was made in isolation, without principle or precedent. In other words, I do not take the maxim of my action to be any sort of law, but only something which I am doing right now. Particularistic willing means, by definition, that the question of whether my maxim can be made into a law which governs me is either absent or is answered in the negative.

So the AAPW goes as follows. To will something, one must identify oneself with a principle on which to act, which would be the same in all relevantly similar circumstances. Willing by a principle is how one becomes separable from one’s incentives: if one is moved only by one’s incentives then the agent’s will has done no work in making the decision. Since Korsgaard argues that agents must solve the mereological problem in order to exist, and they therefore must be identified “over and above” their incentives, then an autonomous, willing agent does not merely act according to her incentives. (Although, of course, she may will consistently with them in a given case.)⁹ If one were to will particularistically one could not be so distinguished from

⁹ It follows from this reasoning that one’s will cannot be the result of (or equivalent to) one’s motive or desire to act according to a principle, or according to reason in general. This is the position of David Velleman that I mention in chapter three, below.

Although I try to present her argument as persuasively as possible, there is a possible inconsistency in Korsgaard here, since she says that normativity is both “guiding and motivating,” with the latter component presumably moving one to action; but her AAPW seems to require that the motivation is *not* what moves one to action, since it is just one of the cluster (or “mere heap”) of incentives. So, for instance, the motivating component cannot come from an agent’s *want* to unify herself, since that want is just part of the

one's incentives; to will particularistically is to have all of the causal power in the incentives and not be an autonomous agent – that is, to have the will be causally absent from the decision – which is not in fact willing at all. Therefore, all willing must be universal, or else it is not done by an agent (and so it is not willing). As I said, this is intimately bound with the need to solve the mereological problem, which one does by performing actions.

When one does *this*, one “constitutes” oneself “over and above” the “mere heaps” of desires, incentives and so on. This is because “[i]f you fail to [act on a principle] it seems that there is nothing to identify you with (or identify as you), and you are nowhere to be found. It is only in acting that you have an identity, in a far more than metaphorical sense.” (EPF 72) In other words, the mereological problem is solved *only if* a universal principle is willed and acted upon. At this point, let me flag that fact that the AAPW only establishes the latter as being *necessary* for the former (hence, the “only if”), but not as being *sufficient*. Recall that the dilemma I am framing for Korsgaard is between whether or not “bad actions” allow agents to self-constitute. Bad actions are actions based on bad maxims. A bad maxim is “I will take my coat off to warm up” or (for Korsgaard) “I will maximize utility in order to be ethical.” If the AAPW established the sufficiency condition, then *any* universal willing would solve the mereological problem, so long as it was *structurally* correct, meaning in the form of a Kantian maxim. But, unfortunately, the AAPW does not establish this. “All actions are willed universally” is not the same conclusion as “all universal willings result in actions.” In other words, even if willing

cluster on Korsgaard's picture. In my arguments later, I will accept the thesis that one can will inconsistently with one's own *wants* in favour of duty, with the demand that wants are defined in some realistic sense. Korsgaard may be able to remove her inconsistency by describing one's “cluster of incentives” in some other sense, but the way she describes it (as that which presents the act as eligible) leaves the Humean open to interpreting her as inconsistent.

based on a principle is necessary for action, this does not make it sufficient for it. If it is sufficient, then bad actions unify; if it is not, then either bad actions do not unify (and are thus indistinguishable from non-actions), or there is no such thing as “bad action” (that is, it is a contradiction).

The structure of action is thus crucial for Korsgaard to spell out. I take the relationship between action and the two Kantian imperatives – hypothetical and categorical – to be the third and final pillar in Korsgaard's theory. A hypothetical imperative is one that takes the form of a maxim “I will do x to bring about y;” given Korsgaard’s definition of action, all actions take the form of a hypothetical imperative, since they are acts done for some end. The categorical imperative (CI) is (using Korsgaard’s formulation in SC22) the principle of acting only on a maxim that one can will as a universal law.

As I said, Korsgaard wants to come up with norms of action that both guide and motivate us. Further, she develops a constitutive account, which she feels has the greatest force against the skeptic. This means that she is simply describing how action already works: she says that she is seeking “self-knowledge. That is to say... a self-conscious appreciation of what you are and how you work.” (SC 67) So she is attempting to account for the guiding, motivating and descriptive aspects of action, and she tries to capture all of them with the categorical imperative.

Korsgaard claims that there are two aspects of agency – autonomy and efficacy – and that these are captured by the categorical and hypothetical imperatives, respectively. But she argues that the latter imperative is subsumed by the former. (SC 70-72) Essentially, one acts according to CI if one wishes to be autonomous – i.e., if one wishes

for oneself to be the cause of some end. But we must be effective, or else we will fail to be such a cause, and the hypothetical imperative is just a prescription for how to be effective. So if one wishes to play an instrument, and sees one sitting over there, then the hypothetical imperative tells one to pick it up (unless it is a piano) and play in order to realize the end. Autonomy – determining oneself to be the cause of some end – is realized by efficacy, by actually *being* that cause. (This is also related to the AAPW: to be autonomous is to will (and be effective), and to will is (according to the AAPW) to will universally.)

This is where Korsgaard gets the guiding and motivating parts of her theory. If one sets out to do something, then there is a way that thing is done, and since that thing is what one has set out to do, one should be motivated to do it in that way, for that is what it is to do that thing. She argued earlier that action is necessary, that we are condemned to it. Assuming we are motivated to self-constitute (an assumption to which I will return), we should be guided and motivated to act according to CI, because conforming to CI is simply what it means to act (which is the descriptive part). (“Unless we are guided by these [two Kantian imperatives] – unless we are at least trying to conform to them – we are not willing or acting at all.” (SC 81))

So the necessity of action, the argument against particularistic willing and the categorical imperative are supposed thusly to combine to give Korsgaard the normativity she is looking for while maintaining the strongest possible force against the skeptic. Earlier, I introduced a dilemma, the disjuncts of which each characterize a possible interpretation of Korsgaard's theory, along with two other aspects (autonomy and connection with reality) against which her theory can be judged. I believe I am now in a

position to judge it, having captured its essential aspects, and it is to this more critical part of my project that I now turn.

2.4

I mentioned, the dilemma I wish to paint Korsgaard as being in revolves around the possibility of bad actions. Action is both necessary and sufficient for agency according to Korsgaard: agents unify through (all) action (which is the sufficiency condition), and there are no agents without action (the necessity condition). Further, CI provides both a *description* of how action works and a *prescription* for how it is supposed to work.

Specifically, following CI allows the agent to act autonomously and, via CI's Siamese twin, the hypothetical imperative, efficaciously. In what I will henceforth call the “first horn” of the dilemma, the interpretation which does *not* allow for bad action, I will examine Korsgaard by the strict letter, though probably not the spirit, of what her theory literally says.

In particular, I would like to latch onto the descriptive role that CI is supposed to play. As I alluded to earlier, Korsgaard wants it to play this role so that her theory can have the strongest possible force against the skeptic. When the skeptic asks why he should do this or that, or why he should do it this or that *way*, Korsgaard can simply say, “you're already trying for that every time you act; if you do it this way, you'll be better at what you're already trying to do.” Dick, in his dissertation, refers to this blunt “why should I do that?” brand of skeptic as a “naive skeptic.” He also describes a more sophisticated kind of skeptic who appeals to a Humean belief-desire model. This skeptic

can admit that, yes, a theory might properly characterize what the right or proper or correct thing to do is, even in a wholly objective sense, but at best it can only lead one to have *beliefs* about what is right; there is no reason to suppose that these beliefs will translate into corresponding desires to bring this right state of affairs about.¹⁰ This skeptic is supposed to be silenced by the necessity of action, along with the descriptive account of what that means; together, these two things yield, “you must do this and this is how you do it.” Korsgaard means to respond to both of these skeptics, so the descriptive arm of her constitutivist picture must be such that she can do this.

So to be an agent is to perform actions, to perform actions is to abide by CI, to abide by CI is to be autonomous and efficacious, and to be efficacious is to conform to the rational dictates of the hypothetical imperative, which tells one how to take the means to bring about a certain end. Therefore, there are no failures of rationality – that is, none *by agents* – of taking incorrect or improper means. Further, since there is necessarily action, in that we cannot help but act, then we cannot help but be autonomous and efficacious in any given situation where we are faced with a choice. If we were to fail at any of this, there would be no agent to speak of who is doing the failing; the fact of the existence of the agent entails the success of the agent at *being* an agent, and to achieve this success is to act in conformance with CI.

If the previous paragraph did not seem at least a little bit ridiculous, then you probably did not read it correctly. I, you and Korsgaard should all agree that it is. It is ridiculous for two main reasons. Most obviously, it is a direct contradiction of one of her stated goals, which is to allow for bad actions. But secondly, what kind of a picture of

¹⁰ See EPF 63ff for the discussion of these two kinds of skepticism. Also, for the “sophisticated” brand of skepticism, see Michael Smith (1994), The Moral Problem.

agents and agency does this give? Korsgaard makes her definition of action narrower than the colloquial one by distinguishing actions from mere “acts,” as I outlined above. And she *further* narrows it – or so I am supposing on this first horn of her dilemma – to only acts which are based on principles which conform to CI. And since action is necessary and sufficient for agency, specifying action by such narrow conditions lessens the frequency of actions and therefore makes it less likely that anything will count as an agent. But what do we call those things which seem like actions but are not, and what do we call those things which seem like agents but are not? This leads back to the concern of how a theory tethers itself to reality. Surely these non-agents exist, and these non-actions occur, but what are we to make of them? And, most importantly, how do these more technical, theory-laden conceptions of these concepts figure into her more substantive claims, such as that action is necessary?

In his essay “Agency, Shmagency”,¹¹ David Enoch takes up the point of using these terms in such technical ways. Enoch says, as a foil for a constitutive project like Korsgaard's, “I am perfectly happy being a shmagent – a nonagent who is very similar to agents but who lacks the aim (constitutive of agency but not of shmagency) of self-constitution. I am perfectly happy performing shmactions – nonaction events that are very similar to actions but that lack the aim (constitutive of actions but not of shmactions) of self-constitution.” (AS 179) We can think of “shmactions” as things which fall under the colloquial definition of “actions” but not Korsgaard's more technical and restrictive definition of it, and of “shmagents” as the things which perform them. Korsgaard's mistake, according to Enoch's and my current criticism, is that she equivocates her

¹¹ Enoch, D. (2006) “Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action.” This is hereafter referred to in the text as AS with page numbers.

technical term “action” with the more colloquial term. Depending on which sense one plugs into a given reading of Korsgaard, the claims can be seen as either implausible or inconsequential. In the case of the necessity of action, the *plausibility* comes from taking the term “action” in its colloquial sense, but the *force* comes from taking it in its technical sense. The skeptic, whose assent Korsgaard covets, will probably agree to the thesis with its employment of the former, but not the latter. But if he accepts it in its latter sense, then he could just, as Enoch says, perform “shmactions”.

Much of what I have said about Enoch's criticism will come back upon examination of the second horn of the dilemma, which allows for bad actions. (Note Enoch's phrasing of having the “aim” of self-constitution.) On the first horn, I am using an interpretation of Korsgaard which takes actions and agents very far from their colloquial meanings, and hence the ridiculousness of the above paragraph. However, it is the far-off meanings that seem to be required in order for Korsgaard to convince the skeptic of the normativity which she claims is inherent in action (even if she does not convince him that “actions” in her sense are what he performs). This is because she is trying to explain to him how action works, and not merely how it ought to work, lest the skeptic simply not desire to bring about what he agrees ought to be. In order to get the conclusions she wants about the normativity that results from what is constitutive of action, she needs a conception of action that will do all of this for her. But such a conception surely does not characterize the *only* things which are done by agents, and this leads us back to Enoch. In other words, it is difficult to believe that a conception of action or agency put forth *without* the aim of establishing the normativity of action would resemble the one I am characterizing Korsgaard as having on this first horn.

Most likely for this reason, as well as to permit bad actions, Korsgaard does not interpret herself in this way. The reason I was steered toward that interpretation was twofold. The first reason was the response to the skeptic which I mentioned. The second reason is the following. Korsgaard gives a functional definition of what an agent is – i.e., one who performs actions. Latching onto the descriptive part of her project, I took conformance with CI as a *description of what it means* to act. If this is what it means to perform an action – if this is what is *constitutive* of action – then things which do not conform to the description are simply not actions. If x is constitutive of being a y, then if A lacks x, A is not a y. To admit of “bad actions” seems to be to deny this – unless they are defined to be anything which is not an action, in which case acts based on bad maxims are “bad actions” the same way refrigerators and the theory of gravity are bad actions. As I said, this keeps to the letter but not the spirit of what Korsgaard says. Her dilemma, as I have been calling it, is between whether or not we rigidly adhere to her definitions. The problem on the first horn is that we find ourselves either rejecting the arguments which employ the terms so defined, or simply rejecting the definitions, which renders the arguments toothless.

“[T]he hypothetical and categorical imperatives are constitutive principles of volition and action. Unless we are guided by these principles – unless we are at least trying to conform to them – we are not willing or acting at all.” (SC 81) To be a bit more charitable, we might seize on one word here, which is *trying*. On the first horn of the dilemma there were two categories: (good) actions and non-actions. On the second horn, to which I now turn, there are three: good actions, bad actions and non-actions. To get bad actions, we need a way to include some of what on the first horn were non-actions

into the category of actions. Taking a cue from Korsgaard, we can tentatively posit that bad actions are those for which conformance with CI is an aim but not an achievement; good actions are those for which CI is aimed at and achieved; and non-actions, as Korsgaard says directly in the quotation, are those for which CI is entirely absent (along with refrigerators, etc.).

How, then, to parse out this “trying” or “aiming”? One way that can be quickly dismissed is that of building it into the maxim. This might be plausible at first blush since one's maxim seems like it should be a complete description of one's means and ends. So, it might be thought, if one is trying to do or bring about y by doing x , one will adopt the maxim, call it M , of “I will do x to bring about y ”. So, if *trying* to conform to the Kantian imperatives is constitutive of action, then in order to be an action one must also have the maxim, call it M^* , of “I will adopt maxim M in order to conform to the Kantian imperatives.” But there are two reasons to resist this approach. First, both of the skeptics mentioned will be utterly unconvinced by this. It is certainly not an innocuous description of how action works that the skeptic will agree to, and the “sophisticated” skeptic will happily perform “shmaxions” which do not have this M^* hovering over top of them. But second, and more important, is that practical reason, as characterized by the Kantian imperatives, does not work this way. Korsgaard wants the rules of practical reason to work like the rules of logic. When Aristotle outlined the rules for syllogisms, he was simply formalizing what were already known to be good or bad arguments; it is not *because* of those rules that they are good or bad. Korsgaard uses the example of modus ponens: to get B from A and $A \rightarrow B$ we do *not* need the rule “if A and $A \rightarrow B$ then B ,” we simply go from A and $A \rightarrow B$ to B by modus ponens. (Also, if we did require a

rule like this on top of modus ponens, then there would be another rule required on top of *that*, and so on into an infinite regress.) Likewise, Korsgaard wants to say, Kant similarly formalized what is going on in good and bad actions. If we want to bring about *y* in a given circumstance *C*, then we simply go from *y* and *C* to a prescription for doing *X* by the Kantian imperatives, in a way that exactly parallels the logic case.

So we do not try to conform to the imperatives in the same way we try to achieve some end. Rather, the hypothetical imperative tells us how rationally to go about achieving our own ends; just as modus ponens is silent on whether *A* or $A \rightarrow B$ is true, the hypothetical imperative is silent on whether any given end should actually be achieved. “Trying” to conform to the hypothetical imperative seems like an easy thing to grasp: one does act-*A* to bring about end-*E* because one thinks that act-*A* is an effective (and consistent with one's other preferences) way to bring about end-*E*. This is supposed to be a large part of the intuitive pull of constitutivism. A skeptic would be loath to deny the pull of this imperative; if he does not do act-*A*, then he must not want end-*E*. One might fail due to some rational deficiency – perhaps act-*B* is a more effective way of achieving end-*E* but the agent lacks the capacity to grasp this. In this case, the agent is still *trying* to conform to the imperative, in that she is trying to be effective, so we can easily say that act-*A* is a “bad” (i.e., less good or effective) action, as opposed to being no action; even though it does not conform to the imperative, the agent still tried. But, having reduced the amount of acceptable maxims to those which conform to the hypothetical imperative, the *categorical* imperative then whittles this list down even further. For example, breaking a promise to get what one wants is perfectly acceptable according to the hypothetical imperative because (let's say) it is an effective means to bring about one's

end. But CI does not allow for it. If *trying* to conform to CI is constitutive of action, then is the person being willfully deceitful simply not acting, and therefore not an agent?

Above, the AAPW suggested – though, as I mentioned at the time, did not *demonstrate* – that maxims yield actions. If this is true then the willfully deceitful person is acting, and is therefore self-constituted. If trying to conform to CI is merely trying to be autonomous – to see oneself as a cause and bring oneself into existence by solving the mereological problem – then the willfully deceitful person seems to be doing this at least. But it is more than this, of course; it *has* to be, if Korsgaard is to get the desired prescriptions out of it. (Once again, the plausibility comes from one interpretation and the force from another.) But I have dealt with the problem of seemingly obvious actions not being counted as actions on Korsgaard's account when I discussed her first horn. Here, on the second, I will allow that any instance of universal willing counts as an action. “Trying” to conform to the imperatives then simply means trying to be autonomous, effective and universal.

But if this is true, then one can not only try but also *succeed* at being autonomous, effective and universal but still commit actions that are “bad” because they do not conform to CI – such is the case with the willful promise-breaker. Is this the kind of account of bad action that is desirable for Korsgaard? The willful promise-breaker – or the willful murderer, or the willful doer of whatever bad action you like – has committed an action in that he was being (or trying to be) autonomous, effective and universal, but he did a *bad* action because the maxim he adopted did not pass the CI test. But if this is true, where does the normativity, either in the sense of guiding or gripping him, come from? He has sought to bring about an end (getting what he wants) and chosen an

effective means to attaining that end (breaking a promise), so where does the grip or authority come from that says he *ought* to have acted otherwise? Or, in other words, he has already asserted himself as an agent in his willful promise-breaking, and has thusly constituted himself, so what, as an agent, need he do further?

2.5

For her part, Korsgaard devotes a chapter (chapter 8, SC 159ff) to trying to account for “bad” or “defective” action. There are two quotations which are indicative of just how she tries to account for them. The first is the following:

[Bad action] is action, because it is chosen in accordance with the exercise of a principle by which the agent rules himself and under whose rule he is – in a sense – constitutionally unified. It is bad, because it is not reason's own principle, it does not rule for the good of the soul as a whole, and therefore the unity it produces... is contingent and unstable. The agent's unity is propped, so to speak, by the fact that the circumstances that would reveal the competing factions in his soul and undercut his efficacy don't happen to occur. (SC 175)

Korsgaard makes the argument that there are degrees of action (and therefore of agency), determined by how well the action unifies its agent, with those actions that unify best being those with maxims passing the CI test. Now, passing the CI test is *not* a matter of degrees, and presumably all actions with maxims passing the CI test are *good* actions. So for Korsgaard, the degree of badness of actions must come from how well they unify and not how well they pass the CI test.

This could be taken two ways. The first is on the level of individual actions: given the (contingent) circumstance C, action A, with maxim M, unifies an agent, since, even though M does not pass the CI test, C is such that the agent's efficacy is not undercut. In this case, it seems that action A unifies the agent perfectly well, since the agent is in circumstance C. The second way of taking it is on the level of actions over time: action A unifies the agent in circumstance C, but (let's say) not in many other circumstances. (And actions with maxims that pass the CI test unify agents in *any* circumstances.) In this case, action A is not very good at unifying the agent, since it has a limited applicability – limited to a narrow range of circumstances. So there is a distinction between a) how well an action unifies in a given circumstance C, and b) how well an action unifies in some *arbitrary* circumstance – that is, how unifying it will be in different situations. Bad actions on a) are those which do not unify well right now; bad actions on b) are those which do not unify well over time.

But it should be fairly clear b) will not give us an adequate account of bad action. This is because efficacy is highly dependent on contingent circumstances. The action with the maxim "I will do act-A to promote end-E" is only effective if end-E is promoted by A'ing in the agent's particular circumstance. And also, recall from the AAPW, the agent limiting the maxim to her circumstance does not diminish the universality of her principle, since she would will the same in all similar circumstances (as she cannot will particularistically). And further, it seems perfectly appropriate for maxims to be situationally dependent: if an agent wishes to be autonomous and efficacious – that is, to insert herself into the causal order – then different actions will better allow her to do that

in different circumstances. It is only practical for the agent to tailor her actions to her situation.

So we are left analyzing the badness of actions on the level of the individual action in its circumstance. Here, the charges Korsgaard lays that the unification is "contingent and unstable" do not seem to hold up, with the first being unproblematic, and the second only being applicable to bad action on interpretation b) that I mentioned – actions whose maxims have limited applicability. On the level of *individual* actions, actions whose unification is contingent and unstable seem perfectly okay. So on what basis can we criticize these actions as being bad? We thusly are brought back again to the problem of bad action.

The willful promise breaker has achieved his end in an effective way. Yes, he has poisoned the well, so to speak, and drained much of the trust others had in him, and so he has made this tactic unavailable to him in the future. Arguing against his action on this basis is the classic Kantian move: he cannot will this in all relevantly similar circumstances because he has rendered this action *ineffective* in all relevantly similar circumstances by doing it this time. The agent is being imprudent by doing it, to be sure, as he has little to no regard for his future self by doing so – for instance, whether the future self can be trusted. But as far as being autonomous and efficacious *now*, there seems to be no grounds for criticizing him. And next time, when he is trying to get what he wants but his circumstance is such that his trust is lost, acting on the same maxim (breaking a promise to get what he wants) will surely be ineffective – in *that* circumstance. But *now*, we lack any grounds to criticize him, at least assuming that his goal is to unify himself *now* with *this* action.

It is easy to see what Korsgaard is going for: she wants good action to be that which unifies over time, bad action to be that which unifies sometimes but is generally unreliable, and non-action to unify never. And this is what is behind her extended analogy, from Plato, with the different types of governments of a state being analogous to the different ways to govern one's soul (which I have not discussed up to now due to its murkiness). A well-governed state will be one whose government will respond well to different situations and contingencies and keep the state functioning well – and functioning *as* a state. However, Korsgaard's metaphysics of normativity, with which I have been primarily working with here,¹² is such that agents are unified through action, and action is something which renders an agent autonomous and efficacious. And it does so *now*, since that is how we identify the agent over and above the mere heap.

But we might still ask whether one action can unify better than another *now*. Korsgaard answers this question in the following way, with the second of the two quotations that I mentioned:

[T]o the extent that an agent's legislation fails to unify her, and render her the autonomous and efficacious author of her own movements, she is less of an agent, and to the extent that she is less of an agent, the source of her movements must be some force that is working in her or on her. (SC 174)

That is, to the extent that the agent is responsible for the action, she is unified by it. And as we have seen, to the extent that the agent is unified by her action, the action is a good

¹² These are the arguments laid out in chapters 1 through 4 of the book regarding the metaphysics of normativity. In the earlygoing, Korsgaard says that the conclusion she reaches in chapter 4 “is not short of the conclusion I hope to reach in the book in general,” (SC 45) because they lay out its foundational arguments. It has been my strategy to work with these arguments primarily. In this section, it has been to back up what I said earlier, and show that her foundational arguments do not support her account of bad actions which she lays out in chapter 8.

one. So there is no component or aspect which is both bad and an action. An action which is 50% bad is only 50% an action. There is nothing which is 100% an action and (say) 50% bad, making it only half as good as another, better action. An action is only worse than another because it is not as much of an action.

We are thus left with the problem of bad action: either it is a contradictory term, or it unifies an agent well enough so that she has no further impetus to perform good actions. This is the problem of the competing interpretations of the first and second horns which will not go away for Korsgaard, and she cannot have what she wants from both at once – to convince the skeptic and generate the normativity she wants.

2.6

What has just happened is what we might have expected to happen. On the first horn, we allow Korsgaard the definitions she needs to get the results that she would like, but those results do not pertain to “agents” or “actions” in any *real* sense, only in her more technical and restricted sense. On the second, we reclaim the colloquial meanings of the terms she uses and her normative conclusions disappear. If any action unifies the agent, he is just as well unifying himself with bad actions; there is no further pull to do otherwise.

What should be clear is that Korsgaard cannot simultaneously satisfy the skeptic, by allowing him an acceptable amount of freedom to wholeheartedly defy the categorical imperative, and extract normativity merely from the nature of action. “Action” and “agency,” it seems, in any acceptable (read: realistic) sense, are too inert, conceptually, to

generate that kind of normativity. Further, it should also be clear that there is a difficulty involved in allowing for agential fallibility in a constitutivist framework without *presupposing* certain motivations of the agent – that she is *trying* to do (or bring about) this or that merely from the fact that she is acting. It is only against this kind of standard – against what she is trying to do – that we can say that the agent has failed. Now, certainly people try for and fail at all kinds of things. But this by itself is not the problem for Korsgaard. Her problem is the ascription of universality to the supposed motivating force of conforming to the categorical imperative, mistakenly equivocating this with the motivation one (plausibly) has to make one's actions one's own. Although the attempt to account for a universal normativity is an admirable one, and one I will pick up on in the following chapters, the method she uses to ground it does not work, for reasons I have argued. It is only against the presupposition that agents are inherently trying to conform to CI that we can say an agent acted badly. On the other hand, if a constitutivist gives an account of what is constitutive of one's *actions* rather than one's *attempts*, which is what I discussed in the “first horn” of Korsgaard's dilemma, then there does not seem to be room in the theory for bad actions at all.

In my next chapter, I will expand on and clarify some of the issues that have come out of this discussion of Korsgaard. In particular, I will flesh out the positions of the skeptics some more, since they play a fairly large part in the discussion. The force of the skeptical position comes largely from the intuitive appeal of one's ability to act however one likes and for whatever reason. It is therefore important to give an account of this kind of freedom and how it can mix with the concept of normativity. In other words, can a truly free agent feel the pull of a normative force, or assent to the authority of a

normative guide? Once this question is addressed, I will be in a better position to evaluate the empiricist position's grounding of duty to see if it fares any better.

CHAPTER 3 FREEDOM, NORMATIVITY AND SKEPTICISM

3.1

It is time now to lay some theoretical groundwork. The discussion of Korsgaard was meant to be a case study of the ways in which a theory of action tries to solve certain problems and account for certain intuitions or phenomena. As such I was more interested in how these intuitions clashed with one another within the theoretical framework, and less interested in the merits of the framework on its own. For instance, I more or less flatly accepted the thesis that “action is self-constitution” for the sake of proceeding with the discussion, and likewise (pretty much) with the argument against particularistic willing. A reader pulling his or her hair out over these acceptances can be forgiven, but they were necessary in order for there to be a gateway into the larger discussion, to which I now turn.

A first step in this discussion is the recognition that the intuitions and phenomena Korsgaard wished to account for are not the only ones possible. For example, many would be quite happy to simply deny that the mereological problem is any kind of “problem” at all. Likewise, some might find it perfectly acceptable not to admit of bad actions, actions where one did what one ought not to have done. For example, some philosophers might assert that what deterministically causes one to act and what provides one with normative motivation and guidance are one and the same thing (e.g. a “desire”), and thus bad action is impossible. So it is difficult to say uncontentiously what a theory needs to account for. Arguing against a theory based on, say, the fact that it does not

solve the mereological problem is not *really* an argument against it if the theorist is willing to bite that particular bullet, so long as doing so is not completely absurd. (One or two bitten bullets might be okay.)

In arguing against Korsgaard, I tried to avoid this as much as possible, critiquing the theory more from within than from without. The two exceptions to this were the subjects I introduced of autonomy and connection with reality. The latter seems obvious enough not to warrant much discussion, as it seems to be a basic requirement if a theory is to be *about* anything. Thus Korsgaard's theory (on one interpretation) might well describe some kind of “agents” (in her sense) were they to exist, but there are not many of them walking around the streets of Halifax, so it does not do us much good. That is, she may have successfully given an account of the constitutive requirements of some kind of agency, but she has not convinced the skeptic that he is an agent in her sense, and he therefore remains unaffected by the normative consequences of her theory.

But it is the former subject, autonomy, which needs to be spelled out a bit more, and its connection with normativity is the main subject of the current chapter. I will therefore do the obvious and spell out the concepts of freedom and normativity and see how they go together. But first, I will outline how this all connects with my overall project.

3.2

While Korsgaard's theory was ultimately unsuccessful, the normativity she attempts to account for has something going for it. As I mentioned, she describes normativity as

both guiding and motivating. Further, and in keeping with her Kantian leanings, it must be *categorical*, in the sense of being “unconditional and unyielding.”¹³ This makes sense in the context of her overall project, which is to account for moral and rational normativity. No one is exempt from the demands of ethics, and trying to convince the skeptic that he should be good seems like a worthy enterprise, however difficult. And rational action also seems like it should be the same for everyone since, as we saw with the hypothetical imperative, it is just a logical outgrowth of what one’s ends are and how the world is. Trying to get all of this from the thin concept of agency and its constitutive standards was the theory’s downfall; however, it is to be applauded for its attempt to include all agents, even those most wary of it, in its attempt to ground normativity.

The basic question I will be addressing throughout the remainder of this thesis is the following: what would make someone agree that she *should* do this or that in a given situation? I mean this to be a categorical “should” – not a hypothetical one of the form “I should do A if I am to achieve E.” Korsgaard’s goal was to grip all agents, such that they would be inherently drawn toward the normative authority she described. And, crucially, agents were supposed to *freely assent* to this normative authority – *autonomously*, in the sense emphasized by Kant. Therefore, I am interested in characterizing a feeling of *duty* which one would pursue by one’s own initiative: duty of a free person.

One possibility that should be carefully scrutinized is that rational agents should follow the norm that one should simply do what one likes. As I will be making clearer in the pages ahead, however, one quite often goes against what one *likes* and instead does what one thinks one *ought to do*. This obviously needs to be fleshed out a bit more, but

¹³ This I have taken from Frankfurt’s ANL 136. (See the list of abbreviations.) I elaborate further on his account in chapter 4.

the basic reason for thinking that this alternative norm is plausibly the norm of rationality is that, while doing one's duty is quite often difficult and not in line with one's self-interest, it is something that one does freely. And considering Korsgaard's arguments in NIR, if the proposition "one does what one prefers to do" is a mere triviality – since one's preferences are revealed in one's conduct – then "preference" (and "desire," etc.) is a meaningless term, a placeholder for anything that moves one to action, however varied and disparate the real instantiations of that definition may be. In order for this thesis to be a *real* possibility, what one "likes" needs to be cashed out in some real terms. Just as Korsgaard's "agent" had to correspond in some way to a human being, these "likes" or "wants" or "desires" need to correspond to states of or emotions had by them.

But that is a subject for the next chapter. In the current one, using the lessons learned in the analysis of Korsgaard's account, the goal is to provide a theoretical framework that will allow this kind of normativity to be put into a proper context of human action. Where do our feelings of duty come from? What is needed to convince us – to convince *any* of us – that we have a particular duty? Using the skeptic (whose position I will define more precisely) as the standard bearer for freedom and staunch refusal, these are the questions I wish to answer. My strategy for this chapter is thus to define the skeptic more precisely – which includes characterizing the freedom he practises and defends – and give an account of the kind of reasoning he would approve of using. With this framework, I can then inquire as to what his duties might be.

3.3

First, then, I would like to use an insight from Gideon Rosen to shed some light on the subject of normativity. Rosen (Manuscript)¹⁴ argues that the normativity which is the result of correctness conditions, like the normativity of constitutivism, only gives a kind of normativity in the sense of something being correct or incorrect and has no bearing on its own about what one ought or ought not to do. As Rosen says, in the case of meaning,

[s]uppose that S means that p, and that p is false, so that an assertion of S would be incorrect in our sense. This is a normative claim about the assertion of S. But it is not by itself any consideration against the act. To say that the act would be incorrect does not imply that there is a reason not to perform it, or that we are motivated not to perform it, or anything of the sort. Of course, there are all sorts of reasons not to make false assertions, just as there are all sorts of reasons not to kill people or to cause pain. But the mere fact that an act violates the standard of correctness that partially constitutes it qua assertion would appear to be, in itself, no consideration at all against performing it. (NMAT 17)

Dick also picks up on Rosen's example of playing Mozart's C-major sonata:

[I]f you play C# where a B is indicated by the score, you have played the sonata *incorrectly*. The score is (let's stipulate) the final arbiter of what constitutes correct or incorrect playings of Mozart's sonata, but the score cannot settle the different, and further, question of *whether you should play the sonata correctly or not*. Say I wish to make a musical joke, or conceal my piano skill from someone

¹⁴ Rosen, G. (Manuscript) "Normativity, Meaning and All That" This is hereafter referred to in the text as NMAT with page numbers.

who would be made jealous by it. Now it looks as if I have ample reason to play the sonata *incorrectly*. (EPF 134)

Dick uses Rosen's insight to argue against Korsgaard's constitutivism in the following way. He gives a detailed account of Bishop Joseph Butler's ethics, wherein Butler asserts that God had designed us in such a way that our conscience plays a central role in our decision making. Further, since "God designed us to consult and obey it," this is what we *should* do. (EPF 135) This is an example of a teleological claim leading to a normative claim: we are designed to work a certain way and therefore we *ought* to work that way. Dick describes Korsgaard's account as being structurally similar: she argues that "it is *just true* that we are beings who self-constitute and so we should do so." (EPF 136) Now enter Rosen: even if we grant to Korsgaard that she has provided the correctness conditions for agency, Dick says, this is not necessarily our aim (recall Enoch's "I'm perfectly happy being a shmagent"), and so this does not speak against the non-unifying "action" that we wish to perform. We might very well remain unconvinced that we *should* be agents in her specific sense. It is this sense of "should" that I am working with in this thesis – a feeling of obligation – which is not a result merely of meeting correctness conditions.

From this discussion can be inferred what I will call the first Rosen thesis (RT1):

RT1: given some achievable end E, with a set of correctness conditions for its achievement, an agent *ought* to take a set of actions to conform to these correctness conditions **if** she *ought* to achieve that end successfully.

This seems like a fairly obvious claim, as it is simply a description of how "oughts" flow through the logic of action. The reason for the "if," and the lack of "only if," is to allow

for the possibility that the same correctness conditions can apply to multiple ends. For example, the end of picking up a musical instrument and the end of picking up a wooden object are not the same, but the correctness conditions for both are met by my picking up the violin sitting in front of me. If the weak RT had an “only if” in it, then I ought to pick up the violin only if I ought to achieve *all* ends which are satisfied by picking it up. But the “if” clause seems clear enough: if I ought to achieve some end then I ought to meet the correctness conditions for achieving it.

But we can also go one step further, taking a cue from the example of playing Mozart’s sonata. In that example, playing it correctly was not necessarily one’s aim. One might instead have been trying to play a musical joke, etc. So no feeling of obligation came simply from conformance to the correctness conditions of the sonata. But the fact that one has some aim does seem to bring some normativity with it. Rosen uses the following example. From “Fred is the last dodo” it follows analytically that “If you want to see a dodo you should take steps to see Fred,” where “the validity of the inference comes from its mediat[ion] by substantive principles of rationality connecting means and ends.” (NMAT 5) In other words, the validity comes from the very concept of practical rationality, from what it means to say that one ought rationally to do something. Likewise, one can infer from the structure of Mozart's sonata that if one wished to play it correctly then one ought to play its correct notes. This is in line with the hypothetical imperative that Korsgaard endorsed, which was simply a non-judgmental prescription for how to bring about one's ends (i.e., the “substantive principle” that Rosen was referring to).

From this, the following, second Rosen thesis (RT2) can be inferred:

RT2: given some achievable end E, with a set of correctness conditions for its achievement, an agent *ought* to take a set of actions to conform to these correctness conditions **only if** *her aim* is to achieve that end successfully.

(Note that I have changed the language from “wants” – as in, “you want to see a dodo” – to “aims” for reasons that I will get into below.)

Before accepting RT2, we need to be clear about what kind of obligation as well as what kind of aim we are talking about. With regard to the aim, it is one that we might call “post-reflective,” or what Gary Watson (1975) might say is in line with one’s *values*. It is certainly possible that one can have an aim to retrieve ice cream from the freezer despite one’s being on a diet and valuing one’s health. But this is not an aim that would live up to reflective scrutiny. It is only the kind of aim that *would* live up to it that entails the kind of categorical normativity I am looking for. It is possible that there is a definition of *rational* action that makes RT2 true for rational normativity given any aim – and this is likely what Rosen with the “substantive principles of rationality connecting means and ends” was after – but only post-reflective aims make RT2 true for the kind of categorical normativity I am interested in. The necessity condition in RT1 flowed through any kind of obligation; as Korsgaard tried to show in her constitutivism, if one is not meeting the correctness conditions for achieving something then one is simply not doing that thing. (Her mistake was in attributing the wrong correctness conditions to agency.) But the normativity in RT2, with one’s aims being post-reflective, is specifically the kind of personal feeling of obligation I am working with.¹⁵

¹⁵ The reason for the “only if” rather than “if and only if” in RT2 is that one can (and very often does) have an aim toward something which is merely permissible rather than something which one feels is a duty. Hence, a post-reflective aim is consistent with duty but does not necessarily entail it.

Note just how both RT1 and RT2 address the concerns of the skeptic, by building off *already existing* normativity. RT1 is simply the transference of duty into action, and RT2 amounts to the claim that one's post-reflective aims are consistent with one's perceived duties, whatever those may be (and if in fact there are any). Note also that this is not a mere triviality – that what it *means* to have post-reflective aims is to land on something which is consistent with one's values – but a more substantive, empirically true claim that such level-headed reflection steers one toward consistency with what one sees as duty.

This normativity is thusly *self-governed*, meaning that one takes it upon oneself (though not infallibly) to conform to it. It is only normativity with this quality that can convince the skeptic. Recall from above the sophisticated skeptic, who would not be motivated by even the wholehearted belief that something is morally right. The reason for this is that he saw morality as a kind of externally imposed standard; just as he may not want to play a certain musical piece correctly, he also might not want to do the right thing. Moral norms, for him, are similar to the norms of etiquette, or to Korsgaard's (second horn) norms of agency. The skeptic simply opts out of conformance, which is something of which Korsgaard, whose project was to convince the skeptic of the normative authority she described, would not approve. The normativity I am describing, by contrast, allows the skeptic to freely choose his own ends and values (more on this below), and so it is immune to this criticism. The hard part is, of course, convincing the skeptic that he values any one thing in particular, and the prospect of doing so is something that I will be expanding upon later.

For now, I wish to close this section on normativity with an analogy with game theory, which is to tie into my account (below) of different kinds of freedom. Basically, game theory is one framework that describes how one goes about achieving one's aims. ("Games," strictly speaking, only apply to situations where there are other players whose decisions affect one's outcome, but I will be stretching this a bit without too much consequence.) It is the game one is playing that determines the normativity for one's behaviour. For example, there was an article that came out in 2007 with the title "Checkers is Solved."¹⁶ What this means is that for any given situation in a game of checkers, there is a move (or perhaps multiple moves that are equivalently good) that one *ought* to make if one wishes to win. Of course, winning might not necessarily be one's objective. Just as in the case of Mozart's sonata, one might wish to conceal one's skill and commit a wrong move intentionally, or make some other move for some other reason. But *if* one wishes to win then one can be criticized if one makes a bad move.

In this language, we can say that Korsgaard (on the second horn) is judging agents by the standards of a game they are not playing; what she characterizes as "bad" action could be done by someone who was only trying to be a "shmagent." Just as a bad poker hand can be a great euchre hand, so too a bad action (in her sense) can be a great "shmaction." And in this language of games, the position of the sophisticated skeptic can be put a bit more succinctly: he is simply choosing not to play the game to which the rules and strategies you are outlining for him apply. For example, he might not want to play Korsgaard's "agency game" at all.

¹⁶ Schaeffer, Burch, Björnsson, Kishimoto, Müller, Lake, Lu & Sutphen (2007)

But we can imagine Korsgaard giving a first-horn response to this argument against her, to the effect that we *do* in fact have this as an aim (“unless we are at least trying to conform to [the Kantian imperatives]... we are not willing or acting at all”), and so, *given* this, we *can* be criticised for bad action. Like it or not, this response goes, we are stuck in the “agency game”, in the sense of “agency” specified by Korsgaard, and we are trying to play it well, and with that comes normativity (via RT1), just as if we were trying to play Mozart's sonata correctly or win at checkers. My criticism of this earlier was that it presupposed certain interests of an agent in virtue *only* of the fact she is an agent, while in fact agency is too inert a concept to presuppose such interests. The skeptic was not allowed an acceptable amount of freedom to pursue whichever interests he chose. It is this idea of freedom that I will spell out in the next section, along with a more precise formulation of the skeptical positions.

3.4

John Stuart Mill stressed the importance, in a political context, of an individual having the freedom to pursue “his own good in his own way.”¹⁷ The main distinction I wish to press on is basically between these two aspects of freedom stressed by Mill: “his own good” and “in his own way.” I stipulate, as a matter of definition, that one is “globally free” if and only if one can act however one likes for whatever reason one likes, without restriction. I do not mean here that such freedom requires a kind of omnipotence so that one is *effective* in everything one tries to bring about; rather, I mean that one is free to *try*

¹⁷ Mill, J.S. (1869)

to do, meaning to exert one's will toward, whatever one likes. As a second definition, I stipulate that one is “locally free” if and only if one has a goal one wishes to achieve and is free to choose the way in which one goes about attempting to achieve it. For example, if I aim to satisfy my hunger, I am free, within that, to (attempt to) consume whatever I wish that will satisfy it (assuming it is consumable, etc.). Just to be clear, I mean my definition of global freedom to subsume that of local freedom. That is, one is locally free if and only if one can choose between alternatives that satisfy a certain goal, and one is globally free if and only if one is locally free *and* one is free to choose the goals that one is to satisfy. These definitions can also be couched in the language of “games” that I introduced above. That is, one is locally free if and only if one is free to choose one's moves within a given game, and one is globally free if and only if one is locally free *and* one is free to choose the games that one plays.

Perhaps an explanatory example or two is in order. Say someone is selling his house and asking \$300 000 for it. He gets an offer of this much from someone, but his daughter gives an offer of \$275 000 (which is all she can afford). He wishes to maximize his revenue, and within this it is better (locally) for him to take the higher offer. However, if he is globally free he may sell to whomever he wishes, so long as doing so is in line with the aim he wishes to pursue (otherwise he is deficient in some way). So, for example, he may wish to give his daughter a decent home for her new family, and so on. But if he *only* wants to make more money and he takes the lower offer, then something has gone wrong.

Another example: say that I wish to go camping this weekend. This is a binary aim, meaning that its satisfaction does not admit of degrees – I either go camping or I do

not. Within this aim, I am locally free to go either to Kejimikujik or to Cape Breton, since they satisfy my aim equally. I am globally free since I chose the aim of going camping, and then I am locally free to choose the way in which I go about achieving my aim. Now suppose two further facts in this case. First, suppose that I am quite geographically ignorant and know nothing of the distances to or locations of these places, except that they are both in Nova Scotia. Second, suppose that I do not like to drive further than I need to and I will always choose the closer location that satisfies my other aims if those aims require me to drive somewhere. Now, in my ignorance I am locally free either to go to Keji or to the Cape. But then a friend or map or helpful website informs me of the locations of these places, relieving me of my ignorance. Given my preferences, I now have *fewer* (local) options open to me.

It is worth noting the significance of this phenomenon. With an aim or preference that can be satisfied to greater or lesser degrees, new knowledge tends to decrease the number of attractive options available. This is not *always* the case; I could have discovered another camping destination, the distance to which is a negligible difference from the distance to Keji, and in this case I would have more attractive options open to me. Likewise, the house seller (who prioritizes his daughter primarily and his sale price secondarily) might have another daughter that offered him \$275 000 exactly, which would increase his options as well. But it seems that these cases, in which the different options line up with the preferences *exactly*, are the exception and not the rule. Most likely, the seller's other daughter will be able to afford a little more or less (or he might like her a little more or less), or the alternate campsite will be a little closer or further away.

With binary preferences, new knowledge will usually increase options – say, new options for where to go camping (ignoring driving distance). But given that knowledge tends to decrease the effective local options for non-binary preferences, such as preferences concerning driving distances, what would we say about “how free” the person is with his new knowledge compared to without? We would probably say that he is *more* free, since he is ruling out options with his own preferences. I am freely choosing to avoid driving the further distance, making use of my new knowledge. If I am globally free, the option of going to Cape Breton is still available to me, but my new knowledge has made it a less appealing option. My freedom to satisfy my preferences is increased by my finding out how to satisfy them best. If you will forgive the phrasing: the truth, it seems, has made me freer, illuminating my optimal path and showing me the way to happiness. This is essentially one of the two contrasting historical views of freedom I will be outlining in more detail shortly, and it is exemplified by the biblical passage, “the truth will set you free.” (John 8:32) Truth, on this view, shows us The Way, rather than many ways, but this “Way” can only gain our assent by showing us how to satisfy our *own* preferences.

What the conceptions of this type of freedom, in which truth shows one The Way, have in common is that they deny that agents have global freedom, and Korsgaard, on either horn, is indeed an exemplar of this. On the first horn playing the “agency game” is a precondition for our very existence as actors, and on the second we are still seen as acting against our own interest if we do not play it, since we are not “unified” as well as if we do. In either case, the choice of which good to pursue is not one that is made by the agent, rather it is one which is determined by some “larger truth” which is supposed to

point the way toward what we ought to do in a way very much derided by Hume (and the sophisticated skeptic) – that is, this truth is supposed to bestow upon us some duty (or “ought”). For Butler, this truth was the fact of the centrality of conscience in our constitution. For Korsgaard, it is the ideal of agency characterized by the categorical imperative, the deviation from which justifies the labelling of one as a “bad agent.” But as we have seen, being a *good* agent does not bring with it the normativity that the skeptic, in steadfast defense of his global freedom, can agree with; this, as I have intimated, can only come from showing how to satisfy his own preferences, or act in line with his own values. I will return to the two contrasting types of theories of freedom, following a more detailed outline in the next section of the skeptical position and the freedom it demands

As a brief concluding remark on local and global freedom, it should be stressed that if an agent is to freely accept her duty then it must be in service to a good which is freely chosen, meaning that she is globally free to choose it. To convince the skeptic of his duty is to lay the (indisputable) facts out and have him make his choice on his own; he can only be enticed and not coerced. Korsgaard, admirably, attempted to do this, but the prospect of being a “unified agent” was met by the skeptic with an apathetic shrug. But the methodology, at least, seems sound. The question of what could entice the skeptic is thus an important one to which I will return.

3.5

After seeing how they work, it is now time to outline the positions of the skeptic explicitly, and following this a more detailed account of the different types of theories of freedom I referred to above. The goal, ultimately, is to try and end a stalemate, or at least suggest a path toward ending it, between those who staunchly refuse to recognize an objective normative authority and those who are open to accepting it. To satisfy the former camp, it is crucial to present the position of the skeptic as forcefully as possible, because if he can be convinced then anyone can.

The skeptic is one who tries to oppose normativity at every turn. He can be characterized by any of the following three stubborn proclamations:

1) (Naive skeptic) Why should I want to do that?

2) (Libertarian skeptic) Whether or not I want to do that has no bearing on whether or not I should do it.

3) (Sophisticated skeptic) Knowing what a good action is does not give me a desire to do it that way (or at all).

We have seen the sophisticated skeptic in action already, standing vigilant against those who would derive an “ought” from an “is”. So far, he has only been convinced by Rosen's correctness conditions, which build off his existing normative valuations.

The first two skeptics can be seen as asserting their global freedom. The naive skeptic speaks to the question of what can motivate a free agent. Surely it is not the allure (if it can be called that) of being a Korsgaardian agent. But what *can* it be? If agency alone is such an inert concept, what can be said to entice the naive skeptic into

anything, and in particular convince him that he has a duty? – for perhaps he would “prefer the destruction of the whole world to the scratching of [his] finger.” (Hume, 1739, II.iii.3) (Although this seems like a difficult prospect, I will give one example of it in the next chapter.)

The libertarian skeptic does the naive skeptic one better. Even if the naive skeptic can be convinced that he inherently wants (or desires, or is inclined toward...) something, the libertarian skeptic still needs to be convinced further that he *should* do it – that it is his duty (or is consistent with it). He does not accept that he is stuck playing some game or other, regardless of what his wants might be. For example, we might convince the naive skeptic that he is in the “pain avoidance game,” wherein he tries to avoid pain if possible. Pain, obviously, hurts, so this might not be a hard sell for the naive skeptic. But the libertarian skeptic would assert his freedom not to play that game, perhaps driving his own hand into a nail, to use an example from Sartre, as a foolhardy, though decisive, philosophical exercise in free will. And having done so, he would not feel that he has contravened his duty; indeed, were he a preaching libertarian philosopher¹⁸, he would probably feel that he has done exactly as his duty requires.

Recall from above the change in language from Rosen's example to RT2 from “wanting” to “aiming.” A plausible case can be made (though I have not fully made it yet) that the libertarian skeptic, in his demonstration of free will, has done something against his wants, but he surely has not gone against his aims, for his aim was to slam his hand down on the nail. This is why that change in language occurred: in order to bring the libertarian skeptic into the fold. This is because the libertarian skeptic asserts his

¹⁸ I mean “libertarian” in the sense of free will and determinism, as opposed to the political sense.

ability to choose duty *in spite* of wants, which RT2 would not allow if it were phrased in terms of wants rather than aims. Assuming this aim passes (or would pass) reflection, it would, as I argued earlier, be consistent with his accepted duties, though not necessarily with his wants.

To anticipate (a variant of) the Humean retort to the position of the libertarian skeptic, we can consider the concept of a “desire,” rather than a “want.” Surely on some level, this retort goes, the Sartrean self-mutilator had some *desire* to prove a point, and this overtook his ever-present, though defeasible, desire not to feel pain. Likewise, one might do the right thing, in spite of the visceral pull of being selfish or vindictive, because one has some desire to do the right thing or not to go against society's rules, and this pulled harder in the other direction. But we must be careful not to expand the concept of desire beyond its limits of realism, just as we were careful when discussing Korsgaard not to stretch the concept of agency too far. The concept of “agent” is instantiated in human beings, which are real existing things, and likewise, “desire” must correspond to something which is had by the things (people) which purport to have them. If a theoretical term is not consistent with the properties of the things it denotes then it is of little use. This is because if the statement “only desires can motivate” is to be informative, it needs to be non-trivial. If anything that moves one to action – be it a want or a feeling of duty or something else – is classified as a “desire” by definition, then the statement is trivial and uninformative, because “desire” is just a “black box” term for what moves one to action. The statement can be made non-trivial by substituting in a definition for “desire” that is different from the unhelpful functional definition I just mentioned, but the things which do in fact move one to action are too disparate to be

captured by that one term (a want and a duty, for example). So the statement “only desires can motivate,” which characterizes the Humean position, is either analytically true or empirically false, and neither of these interpretations are acceptable ones to characterize a substantive philosophical position.¹⁹

So the libertarian skeptic does not accept that his wants determine what he *should* do, or what his aims ought to be. His wants are motivating, to be sure, and given RT1, they will tend to give some prescription for action which meets their satisfaction. But the libertarian skeptic asserts his ability to exhibit control over them and set his aims otherwise. The naive skeptic was at least prone to a kind of constitutivist argument; if Korsgaard showed that his own wants matched the constitutive aims of “agency” in her sense (which she did not), then he would have accepted her resulting normativity. The libertarian skeptic can *accept* this matching but deny the normativity anyway.

So it should be asked: What can convince the skeptic, meaning the conjunction of all three brands of skepticism I have outlined, of anything at all? If he denies the

¹⁹ Consider as a counterargument a position about belief: if beliefs are about such disparate things, then what unites them under the same umbrella term? However, beliefs *do* seem to be defined in this kind of functional way – they are, roughly speaking, “things we think are true.” That is, the claim “only beliefs can be things (propositions) we think are true” is true analytically. That it is true analytically does not seem, to me, to raise any problems; I have tried to illustrate throughout this paper why the claim “only desires can motivate” *does* in fact raise problems if it is true analytically. This is because when the thesis is employed it is meant to be substantive. So, for example, if a typical person who believes the thesis were confronted with an argument for a motivating belief, he would *not* say “well, that belief is thus a desire, by definition, since only desires can motivate.” Rather, he would say, “beliefs *cannot* motivate; only desires can motivate.” Thus, for this person, desire is supposed to mean something *more* than merely “that which motivates” and the thesis “only desires can motivate” is non-analytic.

Consider also the parallel between “only desires motivate” and the thesis of physicalism that “only physical things exist.” The two theses are structurally identical. When confronted with an argument for Cartesian dualism, the physicalist will *not* say, “you are right, and these mental things are physical by definition, since only physical things exist,” and declare himself a “Cartesian physicalist.” Physicalism is supposed to be a substantive philosophical position because “physical” is supposed to *exclude* certain things (and, e.g., “Cartesian physicalism” is supposed to be a contradiction). The proponent of “only desires can motivate” needs to give an account of desire in the exact same way that the physicalist needs to give an account of physical things, so that his thesis means something.

connection between facts and normativity, as the sophisticated skeptic does, and he denies that his wants determine what he should do, then *what* should he do? If he can indeed go against his wants, on what basis would he ever do this? What could possibly give him a sense of duty? Before answering these questions, I will return to the contrasting views of freedom I introduced earlier and place the skeptic in that context.

3.6

Above, I indicated that if a certain preference is taken as given then new information will tend to decrease one's choices, pointing one toward the choice that satisfies one's preferences most fully. As a corollary, if normativity is taken as given, then new knowledge will point to ever-shrinking options about what one should do. One view of freedom uses this type of theoretical setup: one state of affairs (perhaps multiply realizable), at some level of description, is seen as ideal, and the role of knowledge is then to determine how it is to be actualized, in the sense of meeting its correctness conditions, and how it is to be achieved, in the sense of means to its end. New knowledge brings this ideal into focus and sloughs off the bad alternatives. The other, contrasting view of freedom is that there is no such ideal; rather, the individual sets about determining for herself what she ought to do or achieve.

Dennis Dalton, in his excellent lecture series “Freedom: The Philosophy of Liberation,”²⁰ makes this distinction in rigorous detail. He describes this difference in theories of freedom as being “truth possessed” versus “truth pursued”. (FPL L3) He

²⁰ Dalton, D. (1994) This is hereafter referred to in the text as FPL with lecture numbers.

takes Hegel as being an exemplar of the former in his Philosophy of History (1837). Dalton sees Hegel as seeing history as being the playing out of a “grand drama” of God's plan, with freedom being acting in accordance with that plan.²¹ This should sound a bit backwards to a western individualist's ears, but the consideration of what new knowledge does to one's options might make it a bit more palatable. The hard part is, of course, convincing people of the truth of the factual claims that ground the normativity (i.e., God's plan) – assuming that they then do not deny that normativity does result from them, as the sophisticated skeptic would. But for a truly religious person, being convinced of what God's intentions are would not seem to require further argument to establish that it ought to be brought about. Of course, not everyone is religious. But this is the essence of the truth-possessed view of freedom: some truth is supposed to direct one toward what to do, and in so doing one is set free, in the same way that my newly acquired geographical knowledge freed me from my ignorance and showed me the best option for my camping trip. The particular Truth and how it provides normative force and guidance is what distinguishes one truth-possessed theory from another.

An important aspect of the truth-possessed view of freedom is what the Truth sets us free *from*. Of particular importance is freedom from ignorance, as well as from the passions and bodily impulses. In sharp contrast to other philosophers, truth-possessed theorists tend to assert that one has control – particularly *rational* control – over one's appetites, impulses and so on. Hegel recognized that there is a type of “freedom” involved in pursuing one's impulses and the like, which he called “negative freedom,”

²¹ The accounts that I give of Hegel throughout are more to illustrate my points about freedom (and truth-possessed versus truth-pursued) and less to characterize Hegel accurately. Any Hegel scholars who feel that he can escape any overt or implied criticism I make of him should not feel alienated from the conversation, and are welcome to substitute the name “Shmagel” wherever Hegel's appears.

and identified it with “caprice or avarice.” (FPL L3) But for him, there needed to be more, a purpose toward which one's efforts can and should be directed. There is a reason, perhaps, that we are able to exert control over ourselves rather than simply giving in to every impulse, a reason that gives us some normative guidance.

The “truth pursued” view of freedom is expressed in its typical form by John Locke as well as John Stuart Mill (see FPL L2 and L4). Mill, as we have seen, emphasized one's freedom to pursue “[one’s] own good in [one’s] own way.” He was very skeptical of those who claimed to possess the Truth and used it to dictate what others should do with their lives. The truth, rather, is something about which one makes up one’s *own* mind. Unless one is doing harm to others, Mill thought, one should be left alone to do as one pleases. And Dalton emphasizes the importance of the connection between this and Mill's famous view of truth as being illuminated by its “collision with error.” (FPL L4) Truth, for Mill, is a continuous process of discovery and experimentation where the proper attitude to have is humility. When confronted by someone with a view that is seemingly wrong, we can only try to convince and not coerce him; if his view is wrong then this fact will be demonstrated through the process of rational discourse and the confrontation between the wrong view and the truth which contradicts it. In this way, truths are always being refined, sharpened and strengthened. Mill therefore accepted and encouraged pluralism within a society. As we know very well in our contemporary society, people have a wide range of values, and therefore different ideas of what “the good” is (if they have an idea of that concept at all) – hence Mill's assertion that a person should be free to pursue “his own good in his own way.”

One's "good" will be based on one's *own* values (using Watson's term again), which others should confront with reason and humility.

A key tenet in truth-pursued theories is the willingness of a reasonable person to change her mind based on good arguments. In fact, both truth-pursued *and* truth-possessed theorists might accept this, except the latter would assume that all reasonable minds will converge on the truth(s) they espouse. (It is the insecurity in such truths that lead totalitarian states to be so brutal; people need to be coerced because they are not convinced.) But the truth-possessed camp might also, depending on the theory in question, be caught in a theoretical trap from which there is no escape. What I mean by this can be seen from the following two examples:

Take a (particularly bad) paranoid person who thinks, "They're all out to get me." When people try to reason with him, for example, by pointing out the seemingly benevolent acts people do, he responds by asserting that they are merely trying to earn his trust before they ultimately betray him. He sees all of the acts of others through this lens, without exception, with nice acts being thought deceitful and bad acts showing people's true colours. The key point here is that there is, *in principle*, no observation or reasonable argument that can make him budge from his position. He may decide to abandon it for *pragmatic* reasons – perhaps it is too psychologically taxing to be so untrusting, and so he is better off taking people at their word – but this is not the kind of argument that would really *convince* him, in the sense of rationally changing his mind. Or take the case of a devoutly religious person, who sees the world as God's creation, perhaps with Him even taking an active role in its operation. Everything good that happens, according to this person, is a manifestation of His love, and everything bad

either serves some greater good (which may require a long time to be discernible) or is a test of faith or is a result of His “mysterious ways,” etc. Again, there is literally no piece of information that can displace this person from her worldview (though there can still be pragmatic arguments and the like).

What these theoretical traps have in common is the ascription of properties to the unknown to make an assumption consistent with that which is known. For the paranoid person, the unknown was the content of others’ minds – their real motives. For the religious person, the unknown was the totality of the overall “plan,” since if this were known then the goodness of one part of the plan could be objectively judged (the Euthyphro problem notwithstanding) by how it fits in with the rest of it.²²

So what is the significance of these theoretical traps for the different views of freedom? For Mill (and the like), rather than error being exposed as such by its “collision” with truth, it is possible that the two will just keep colliding with neither emerging the victor. For Mill this would be okay in one sense, since humility dictates that we should allow people their opposing viewpoints; but in another sense, it takes much of the power from the marketplace of ideas as an engine for determining truth – although maybe multiple interpretations, each consistent with known data, is the best we can hope for. For Hegel (and the like), this means that there is no guarantee that whatever truth is being argued for will gain acceptance merely by reasoning with people.

²² I am assuming that, unlike Bertrand Russell’s “lunatic who believes that he is a poached egg,” the people in my two examples can in principle be reasoned with. However, the key difference lies in their projection onto the unknown of whichever properties support their desired narrative. Bayes’ theorem (on one interpretation), says that an individual should lower his subjective belief in something if new evidence makes it less probable. But in my two examples, the new evidence is *just as one would expect*, given the theses brought to the table. (“You *would* say that, wouldn’t you!” says the paranoid person.) By contrast, Russell’s lunatic has evidence available which actually goes against his thesis (assuming normal perceptual capacities and so on). (The Russell example is from Russell, B. (1946), p. 645.)

And since the normative conclusions that stem from these truths are only (or, at most) on as firm a ground as the truths on which they rely, a truth-possessed theorist is not likely to bring many into his fold using reasonable argumentation alone, though a little rhetoric or coercion might work.

Returning to the skeptic, we can see that he would not be convinced by a truth-possessed argument if he thinks that its proponent is simply caught in a theoretical trap. “Yes,” he might say to an ardent Hegelian, “if you look at it that way, I should indeed do as you prescribe. But when examined in *this* way, there is no such prescription. At best, you have convinced me of one possible normative avenue, but it is one among several. Epistemically, you are no better off than this other view here, which allows me to do as I please.” (I am being deliberately vague about Hegel's account in particular, being interested more in its structure than its finer details.)

And yet, it is only in the context of a “truth possessed” type argument that the skeptic will be convinced that he has some duty. He will approach all supposed truths with a good deal of reticence – and, well, skepticism – but he must accept at least one of them to voluntarily go against his own inclinations (if that is in fact what his duty requires). This is because, in doing so, he recognizes something as being important, or at least as being important *to him* – but in any case, important *for itself*. It is only in accepting the truth of that importance that he can be swayed. So while the skeptic is strongly attracted to the truth-pursued view, it is only in accepting *some* truth that he will be convinced to do his duty (and it is my goal, next chapter, to show that he can be so convinced).

So in the context of these different views of freedom, the skeptic employs something like the following skeptical algorithm when someone is trying to convince him that he has some duty. First, he scrutinizes the truth being appealed to which supposedly grounds the normativity, seeing whether it is consistent with what he believes, or at least with what he believes and is not willing to give up. For someone with such truth-pursued leanings, this is heavy scrutiny indeed, for he is not *looking* for a “final truth,” so he must be decisively convinced of its finality. Second, if it lives up to this level of scrutiny, he questions whether its proponent is in a theoretical trap, looking for an epistemic alternative to it which does not have that normative baggage. Third, if there is no such alternative, he makes sure that the normative pull and guidance do indeed follow from the truth he has now accepted, in the manner of the sophisticated skeptic. And fourth, perhaps in his skeptical death throes, he asserts his global freedom to choose his own ends, regardless of which truths are appealed to by others. Therefore, to pass this fourth stage, the would-be persuader has to convince the skeptic that this *is* the end he would choose on his own, thus allowing the skeptic to keep his global freedom.

And now the question comes up again of just what *would* make the skeptic choose this or that end on his own. We have already seen the problems with grounding this answer in the nature of action and agency. Mainly, these concepts were either too thin to be prescriptive or too thick to be convincing, depending on the conceptions being examined. The skeptic will resist the latter in favour of some alternative on an equal (or better) epistemic footing – “shmagency,” perhaps. So it is natural to wonder what it is the skeptic uses this hard-fought freedom *for*, exactly. The most suitable position to advance toward the skeptic, one might think, is the idea that he simply ought to do

whatever he wants – that, yes, he has the freedom to do otherwise, but doing so requires him to value something over his own wants, and he seems to oppose this idea at every turn. *If* there were some truth he could accept, along with the normative conclusions therefrom, then he could make use of this freedom of his; however, there do not seem to be any that can run his skeptical gauntlet, so his wants and their correctness conditions are all that remain. If you do not love any of the available bachelors, then you might as well choose the most handsome.

This, I take it, is one interpretation of the empiricist position. This position either denies one's feeling of personal duty, or couches it in terms of one's desires. But as I have indicated, doing this either stretches that concept to the point of being unrealistic, or it fails to account for the fact that we act against our own interests in the name of duty quite often. In the next chapter, I will expand on this more fully. And following this, I will give an example of an empiricist project which would convince the skeptic in the way I have indicated: by giving him a truth with which he would agree, which leads to duty to which he would freely assent.

But first, a brief recap of what has been done in this chapter is in order. First, I outlined the conception of normativity of which the skeptic needs to be convinced. Essentially, what needs to be addressed is whether the skeptic can be convinced that he has some duty, something for which he would go against his self-interest. The skeptic, as I defined him, can be seen as asserting his global freedom, especially the part about choosing his own ends. His freedom entails that he can simply choose not to conform to the correctness conditions of being a certain way, so any appeal to such conditions must be such that he would choose conformance on his own. It is thus crucial to illustrate what

would make him choose to conform to a certain standard. If there is such a standard to which he would conform, then acquiring new truth will bring this into focus and bestow a duty upon the skeptic, one which is freely chosen and agreed to.

Throughout, I have been painting a picture of the skeptic as someone who has contingent wants, feelings, urges, emotions, biases and so on. Thus, anything that can be *categorically* ascribed to the skeptic will probably be at a more general level; we probably could not convince the skeptic that he has a duty to eat chocolate cake, for instance. The idea that duty comes from one's own preferences is problematic, but it is at a suitable level of generality. Now it is time to see if there is a suitable empiricist project, one which can ground the duty of the skeptic in empirical fact.

CHAPTER 4 THE EMPIRICIST ALTERNATIVE

4.1

In chapter 2, I critiqued Korsgaard's theory and suggested that it was a “Kantian” theory. This is not, or not just, because Kant was an intellectual inspiration and precursor to Korsgaard, although it is clear from her book that he is. Rather, it is because she follows Kant in being in a “rationalist” tradition of action. What this means is that prescription for action comes from an action's being logically consistent, which, for Kant and Korsgaard (and others), comes from its passing the CI test. The demand for passing this test, as we saw, was something of which the skeptic could simply opt out. Korsgaard's attempt at a response to this was that an agent could not wholeheartedly opt out in this way while maintaining her agency, but the obvious empirical refutation of this is that agents, in fact, do.

The rationalist tradition is strongly resonant with what I called in the previous chapter the “truth-possessed” view of freedom, which is the view that freedom involves the convergence to the best or ideal state of affairs (or sets of such states) to which the truth guides us. The rationalist tradition, in general, attempts to make use of philosophical – logical or metaphysical – truths and apply them to the world. (“Nihil ex nihilo” is one such example.) If any such truths are *evaluative* – such as “something is better than nothing” or “rational action is better than irrational action” – then it is not a far leap toward the belief that a free person should assent to and try to actualize that which the truths prescribe. (Gleaning concrete prescriptions from them is another matter entirely.)

Being in the rationalist tradition of action thus seems to be sufficient for being in the truth-possessed tradition. However, it is not necessary. This is because the truths being appealed to need not be philosophical in nature; they might instead very well be empirical. I mean “empirical” here in a very broad sense. Hume, who is quite rightly named as one of the great thinkers in the empiricist tradition, did much of his work from the armchair, looking inward and honestly assessing the truths he found there. So among what I mean by “empirical” truths are those which are phenomenological, emotional, dispositional, and so on, along with the more common-sense denotations of the term – observable, quantifiable, etc. So there may be an empirical truth-possessed programme as well: if empirical truths can point to “better” states of affairs – at least from the perspective of the agents involved – then free agents will, recognizing this, be drawn to these superior states.

There is a longstanding tradition of opposing this as a possibility, a tradition of which Hume himself is a quintessential example; and I have taken up this concern by way of the sophisticated skeptic. This opposition refuses to accept normative conclusions from non-normative premises – hence, the famous “is/ought” gap. But the “ought” I am working with here is dissimilar enough from the one in the “is/ought” gap to not only avoid the problem of traversing the gap but also find some common ground with those who say it cannot be done. This is because the methodology I wish to employ is, in broad strokes, the same one as Hume’s, as well as Adam Smith’s. Smith spoke of what he called the “moral sentiments,” such as one’s desire for praiseworthiness. Now, I am not interested here in morality per se, only in one’s personal feeling of duty – although, later, it might be interesting to give some concluding remarks about the former’s connection to

the latter. But so far as one's own personal duty is concerned, the moral sentiments of Smith and Hume do seem to bestow it or something like it. There will be some disagreement between us as to whether such sentiments need to be couched in terms of one's "desires," for reasons to which I alluded earlier, but the basic idea of the moral sentiments fits in nicely with my overall project. What they give is an empirical fact of one's own feeling of duty – a duty toward beneficence, for example. Once this duty is felt by the agent, no matter its genesis, then the transduction of duty into action occurs in the way described in the previous chapter: the agent freely converges on a duty's prescriptions and attempts to meet its correctness conditions by setting his aims on them.

So the moral sentiments give a type of example of a guiding and motivating duty to which a free agent might assent, and it is this type of empirical project which is the subject of the current chapter. The problem with the rationalist picture is that the philosophical truths on which it depended did not translate into duties that were genuinely felt. The empiricist project, by contrast, as I am outlining it, avoids the problem posed by the sophisticated skeptic. Rather than try to bridge the "is/ought" gap by appealing to the normative demands of philosophical truths, this approach simply appeals to duties which are already there (such as those entailed by the moral sentiments, if in fact the sentiments can be shown to be present and the duties can be shown to be entailed by them). It is thus a project geared toward self-knowledge and integrity (two phrases which figure prominently in Korsgaard's book), toward getting one to have the courage of one's own convictions. By discovering what these are, one will be better equipped to freely act in line with them.

Of course, this project hinges on the possibility that there are such truths that apply across the board. The rationalist does not face this concern, since logic and reason are (we shall grant) universal. But the empiricist is more constrained to what he can see, although certainly he can make abstractions from that. Hume has some (problematic) passages where, for instance, he speculates on the sentiments of “Indian” culture which were different from those of his own, but he then tries to work toward more fundamental ones that span across cultures. Even if they did, there is no guarantee either that they apply to all agents qua agents, or that they apply to all individual members of the cultures they span. To get a kind of categorical normativity, it seems we need at least one of these. I think it is correct that only the rationalist approach can get us the first. That is, in order for a duty to apply to an agent *solely* in virtue of her agency, it must come *necessarily* from the very fact of her being an agent. Necessity, Hume reminds us, is not “in the data” of our experience. So again, we should look to Hume and Smith for the appropriate level of generality for the empiricist project: general enough to apply to agents as we know them, but not applying to all agents necessarily. As we saw earlier, agency is too thin a concept to generate any kind of normativity, so this should not at all be surprising.

It should also not be surprising that the skeptic can only be convinced by a duty which is categorical in this way, for if it were not he could simply take the position of someone to whom it does not apply, which has been his strategy all along. Dick calls this the “categoricity requirement” (EPF 63ff), and it comes from Korsgaard's particular aim of grounding *ethical* normativity; if the rules of ethics apply to all agents, then so too must Korsgaard's account of normativity. But I am approaching it from the other

direction, asking whether there is such categorical normativity *first*; following this, one might *then* ask the question of what kind of normativity it is, ethical or otherwise. The first part of the project is therefore more anthropological or psychological than philosophical: it questions whether agents actually perceive themselves as having some duty, not as a matter of conceptual necessity (as in the rationalist project), but as a matter of contingent but universal facts about the things which actually are agents (human beings, for example).

The strategy for the empiricist project is thus as follows. First, find an empirical truth, to which any globally free agent would agree, that entails some kind of duty of the agent. This truth will either have a duty already embedded within it, such as “it is true that agents feel a duty toward X,” and prescriptions for action will follow in the manner of RT1; or, the agent will be shown to have some set of values based on her post-reflective aims (RT2) and feel a corresponding duty toward that which is valued. In either case, a free agent will converge on the demands of her duty, in the manner of a truth-possessed freedom. This is a necessary aspect of *any* project attempting to capture a categorical duty, since a truth-pursued theorist, by definition, does not accept truths that are “final” in this regard, always seeking revision, clarification, specification and reinterpretation. The acceptance of one's categorical duty – and certainly the *skeptic's* acceptance of it – precludes any of this from happening. If the skeptic is to converge on any duty, then it must be because of accepting some final truth, even if only in general terms.

4.2

There are two main tasks remaining in this chapter. One of them is to give an idea of how an empiricist project along the lines I have been describing might work, and to do this I will present a variation of Harry Frankfurt's idea that we act in the interests of that which we love. Frankfurt presents a compelling case for unconditional (i.e., categorical) duty which is felt on a personal level, but which nonetheless is the result of a contingent empirical fact. I take his to be a paradigm example of the kind of duty to which I have been referring.

The other task is to tie off a loose thread from earlier with regard to the separation of duty from desire. Since the latter concept often figures quite prominently in empiricist accounts of action, I need to make it clear that I wish to separate it from my concept of duty. To be sure, holding onto the concept of a desire will be compatible with much of what follows if one wishes to contort it enough to make it fit. We might “desire” to be beneficent to others, as Smith thought, or we might “desire” to act in the interest of that which we love. But since we wish to give an *empirical* account of duty, we must make sure that our terms – or at least the ones that play causal roles – have some empirical basis.

And as an empirical concept, desire seems quite vacuous. The reason for its prominence lies, I believe, in a fairly deep philosophical malady. Philosophers have a hard time dealing with phenomena which are of a certain quality coming from or affecting things which are not of that quality. The “mind-body problem” was only a problem because of the difficulty in accepting that mind can come from non-mind, or

consciousness from non-consciousness, or mental from physical, as well as in accepting any causation between them. The free will problem comes from the difficulty in accepting that free things can come from determined (i.e., not free) things. Likewise, the “desire theory” of motivation hinges on the idea that desires need something which is qualitatively similar to be affected – that is, that desires motivate, pull one in a certain direction, and so something of the same type (i.e., another desire) is needed to pull in another direction to counteract or overwhelm that motivation. Just as a mind, according to several mistaken post-Cartesian theorists, could not possibly affect a physical thing, so too a belief (for example) cannot affect one's motivation, since it is different in kind from a desire. That desires motivate is trivial, since as a technical term it is more or less equivalent to “that which motivates.”

For example, Velleman (1992) argues that the “functional role of an agent” must be played by a “desire to act in accordance with reasons,” (p. 479) since only desires can motivate. An agent (or the functional equivalent thereof), on Velleman's view, weighs reasons and “throws her weight behind” whichever ones she deems appropriate. Velleman is, quite rightly, accounting for the real fact that an agent (or some part of her) intervenes when her desires are taking her somewhere she thinks she should not be. But he is mistaken, for reasons I mentioned, that it must be another desire which does this. (Watson (1975) similarly asserts that the values about which he theorizes are a species of desire.)

The mistake lies in the assumption that desires must be *counteracted* if they are to be rendered ineffective. But the truth-possessed tradition, and in particular the

rationalists, teach us that one can exhibit *control* over one's desires. It is not that they need to be counteracted; they can also be *overridden* – by rational control, for example.

In addition to the empirical problems with “desires” that I have mentioned, this connection with the truth-possessed tradition illuminates another key motivation for avoiding the thesis that only desires can motivate, which is the avoidance of Hegel’s “negative freedom.” That is, if only desires can motivate – and if a *realistic* conception of that term is employed in that thesis – then freedom is indeed identical with caprice or avarice. The recognition of one’s duty, and the prioritization of *it* over one’s desires, avoids this. To be free is not simply to be a wanton.

The talk of “overriding” desires might naturally lead one to the conclusion that the mereological problem must be solved. It is natural to suppose that *I* am the one who takes control – and there is good psychological evidence that in fact we do naturally identify ourselves as this kind of rational control-taker; see, for example, Kahneman (2011), for an explanation of this. But weakening this a bit, it might be thought, rather than identifying oneself with this rational function, that unless this function is active my actions cannot fully be attributed *to me*. On this view, we need not equate one's rational self with one's whole self, but we must at least give the rational self a chance to affect the proceedings; that is, its participation is a necessary (but not sufficient) condition for *my action*. This indeed gels with what I referred to as one's “post-reflective” aims, in that the rational part of the self is what does the reflecting.

And further, the requirement for the participation of the rational part of the self for action avoids the oft-cited problem of “alienation.” I have not dealt with this issue in great detail, as the discussion has not demanded it, but Korsgaard’s concern about

twitches and the like from chapter 2, above, is an example of it. Her concern was that twitches and slips of the tongue should not count as actions, since they lack a kind of wholehearted participation of the agent, and to the extent that this participation is not wholehearted, the agent is alienated from her movements. This seems correct, but her mistake was to move from this to the requirement that the mereological problem be solved, in the sense that actions must come from the *whole* agent. This mistake seems to be a result of the equivocation of the rational part of the self (which Kahneman calls “system 2”) and the whole self. A softer way to put it, as I have been suggesting, is to assert only that the rational self’s participation is required for action, allowing that there are other factors which contribute as well.

But this seems consistent with what most empiricists have to say, excepting that they translate such considerations as rational control into “desires” for reasons that I mentioned. In their case, desires are compared with one another in a more or less additive fashion; in mine, it is more complex, but truer (I believe) to the empirical reality, in that disparate phenomena are not lumped into the same category. But in both cases we can allow for an agent to have some measure of control over her action in the face of her desires, and in neither case do we need to equate either the desires or the control to one’s “whole self.” And I believe this is the happy medium between rationalists and empiricists with respect to alienation and rational control, where the disagreement was originally due to a false equivalency by the former, and a category mistake by the latter.

4.3

As an example of the kind of empirically grounded duty that I am looking for, I would like to present Harry Frankfurt's account in "Autonomy, Necessity and Love" as well as in "On Caring."²³ Actually, Frankfurt's view is quite similar to one presented a couple of centuries earlier by Jean-Jacques Rousseau. Dalton presents Rousseau as one of the "truth possessed" theorists about freedom. For Rousseau, the truth being espoused was that freedom involved acting in accordance with the general will. He believed in (and argued for) the basic decency of human beings, but he thought that we are socialized into being greedy and selfish. Rousseau argues for a kind of "moral freedom, which alone makes man truly the master of himself. For the impulse of appetite alone is slavery, and obedience to the law one has prescribed for oneself is freedom."²⁴ As with the other truth-possessed theorists, there is an emphasis on control over oneself. In this case, the control is directed toward our own self-imposed law – or *duty* – to which the strictures of society often blind us. Were we not so blinded, we would, Rousseau argues, take it upon ourselves to act in accordance with the general will, and "every man is virtuous when his particular will is in all things conformable to the general will, and we voluntarily will what is willed by those whom we love."²⁵

²³ These are both in the same collection (Necessity, Volition and Love): Frankfurt, H. (1999). "Autonomy, Necessity and Love" is hereafter referred to in the text as ANL and "On Caring" as OC, each with page numbers from the collection.

²⁴ This quote is from The Social Contract (1762) (I,viii,3).

²⁵ This quote is from A Discourse on Political Economy (1755), Section II.

Setting aside the societal malady that Rousseau mentions,²⁶ his idea of transcending one's appetites to act in accordance with one's *self-determined* duty is exactly along the lines of the kind of duty I am trying to capture. And further, grounding this duty in the interests of that which one loves is exactly the kind of grounding which can grip the skeptic. I hope to make this clear by expanding on the duty and its grounding, and to do this, I will use Frankfurt's superb characterization of one's autonomous action in the name of love.

That agents do in fact love is the contingent but (in principle) empirically verifiable fact on which I wish to ground an agent's personal feeling of duty. (If it was unclear earlier just how some fact, particularly an empirical fact, could entail one's personal duty, I hope this resolves it.) This is only meant to be one example of such grounding, leaving open the possibility for others, but it is one which, crucially, lives up to the skeptical challenges that I have tried to raise as forcefully as possible. Like Korsgaard's "agency," our loving is a fact of our constitution, but it is not a conceptual necessity in virtue only of our agency; rather, it is simply how we happen to be made up.

²⁶ I do not mean to set them aside because I disagree with them. I think the effect of societal factors on decision-making, desires and perceived duties is important and worthy of study. My interest here has been at more of a personal level – i.e., how one perceives one's own duties – without much discussion of the important topic of the genesis or wider legitimacy of those duties. I think Rousseau is quite right that contemporary society (both ours and his) places an emphasis on selfishness. In particular, it tries to convince us that we are unhappy and/or insecure with what we have – that we should want more, that we should hold on more tightly to what we do have, keeping it away from others, and (most abhorrently) that we should feel entitled to what we have because it is the result of "hard work" or some other such nonsense. What this distracts and diverts us from is what we find important, where that term is meant to denote the values which pass our reflective scrutiny (as I discussed in chapter 2). Rousseau captures this by the quote above about acting in accordance with that which we love, and it is Frankfurt's excellent expansion of this idea that I wish to pick up on here. But the question of where our values come from, and *why* we love what we do, is an important one that I do not mean to diminish.

Frankfurt goes to great pains to separate his view of love, and acting in the interest of one's beloved, from what Kant called heteronomous action and, I think, he succeeds. Heteronomous action is action in which the will is not a full participant; or in other words, it is action in which the agent is alienated from his or her act. When the will is so alienated, one is capricious instead of free, selfish behaviour almost inevitably ensues, and one frequently ignores one's duty. But love, argues Frankfurt, is not like this. It is a selfless²⁷ activity of the autonomous self. The lover's "primary goal is not to receive benefits but to provide them. He is motivated by an interest in serving the interests and ends of his beloved rather than by an interest in serving his own." (ANL 133)

This *lack* of self-interest is characteristic of love (except see the previous footnote), and, Frankfurt says, it shares this quality with moral duty. "In cases of both sorts – those involving love and those involving [moral] duty – it seems to us that we are not free simply to do as we please or as we wish; love and duty alike generate in us a sense that we have no choice but to do what they require." (OC 170) For Kant, moral duty is the paradigmatic sufficient condition for autonomy because in abiding by it the will is *free from* (there is that phrasing again) its surrounding contingencies, such as the appetites. However, Frankfurt argues that

despite the fact that unconditional love is a personal matter, what a person loves may be among his essential volitional characteristics. What autonomy requires is

²⁷ Frankfurt is careful in "On Caring" about using the word "selfless" in this way to describe love (OC 167ff). This largely stems from his requirement that one can love oneself, in which case "selfless love" is contradictory. However, for now, it gives a good feel of how love works vis-à-vis the will.

not that the essential nature of the will be a priori, but that the imperatives deriving from it carry genuine *authority*. (ANL 135)

Thus, on Frankfurt's view, acts of love are done autonomously, in free submission to (and thus legitimization of) the authority of the interest of the beloved. And the imperatives of love, "despite their manifest contingency, may be quite categorical." (ANL 130) This categorical submission to a normative authority describes what I – though not Frankfurt – have been calling one's personal feeling of duty. (I believe this is because Frankfurt reserves "duty" specifically for moral duty. However, in everyday parlance, it seems correct to say that one feels a duty toward one's beloved, and I think he would agree with the general sentiment of what I am saying.) So, given that love can ground this kind of duty, it is time to see if the skeptic will agree with it.

First, the naive skeptic: "Why would I want to do that?" he asks about an act which is in the interest of his beloved. "Do you not love this thing?" comes a response. "And does that not mean that you wish to see its interests fulfilled? You are free to wish otherwise, but if you do so and still claim to love it then you are clearly deficient in some way. Wanting to see its interests fulfilled is part of what it means to love something." To this, I submit, the naive skeptic concedes.

The libertarian skeptic might then wish to weigh in. His is the challenge that he is globally free and therefore free to choose his own ends. With regard to acting in his loved ones' interests, he could assert proudly that he is not going to play that game, and in doing this he watches bad things happen all around to his loved ones that he could have prevented if he had not decided to participate in this protest. Being globally free, it is of course possible that he does this, but he does it, I submit, in spite of himself (just as

the camper could choose to drive a longer distance but would not enjoy doing so). He feels the grip of acting for the benefit of his loved ones but he exercises his freedom to resist it, and since this grip comes from a feeling of *love*, he is neglecting what he experiences as a duty. It is not merely that what he is doing is unpleasant; he is acting against his own conscience. His freedom, therefore, without the strictures of heteronomy that Kant spoke of, will point him in the direction of his duty, and his love is such that he cannot simply opt out as he has done before.

Last is the sophisticated skeptic. “Yes,” he says, “I may love something, but why does that give me a duty?” But again, it is the *personal feeling of duty* that I am after. It is not the mere fact of his love which is supposed to convince him, but the love itself; *loving* something is different from *knowing that you love* it. His feeling of love will already give him a feeling of duty; there is no further work to be done by informing him of his love. In this way, he is not deriving his duty (an “ought”) from the fact of his love (an “is”). Rather, he feels some duty as a result of his loving something; he is not deriving it from the fact of his love, but from the love itself.

Returning to Rousseau, the sophisticated skeptic (who basically represents Hume) would be skeptical that acting as one would in a state of nature – i.e., Rousseau’s “noble savage” – is how one *should* act. But it is not merely the *de dicto* fact that it is humankind’s “state of nature” that makes it how one should act; the Hobbesian should be lurking around the corner to point out the absurdity of this. Rather, it is because that state is meant to represent one’s *free self*, and it is society which restricts one’s freedom in important ways. (Hence Rousseau’s famous line, “man is born free and everywhere he is in chains.”) That is, *de re*, the state of nature is one in which people are free from the bad

influences of society, free to determine themselves. Whether this is the actual “state of nature” without society is really immaterial; rather, it is an idealized version of freedom, of one’s rational, controlling self, of one’s *autonomous* self. And in this state, as Frankfurt and I have argued, what one loves is of vital importance. The sophisticated skeptic arguing against this state of nature *de re* effectively amounts to him arguing against his own autonomy (which seems a bit absurd). And if he chooses to go against its prescriptions he will, again, look on as his loved ones suffer.

While it is a contingent fact that one has love for this or that thing, it is an *analytic* fact that love cannot be apathetic; this is how a Humean can feel a sense of duty.

4.4

Earlier, I remarked when discussing Hegel that convincing a truly religious person of the fact of God’s plan did not seem to require a further argument for its normativity; for the religious person, it would be obvious that such a plan *should* be actualized. But not all people are religious (and thankfully, fewer and fewer are every day). The goal of this chapter has been to answer the question of what all agents in fact *are* such that they would feel some duty. The lesson of chapter 2 was to avoid looking for a conceptual answer, an answer purportedly sourced in what agents are in virtue solely of their agency. Rather, the goal was to ground this duty in an *empirical* fact of our constitution. That we are creatures who love is contingent; that loving creatures feel duty is necessary.

I have not meant the consideration of love to be exhaustive with respect to grounding one’s duty. Rather, I meant to present it as a paradigmatic example of a

consideration that would convince the skeptic. If something like Hume and Smith's moral sentiments can be shown to be as pervasive and universal as love (and, again, are not couched in terms of "desires"), to the point where the skeptic would be convinced, then the duties they ground would similarly follow. And whether or not this can be done in the case of Hume and Smith and the moral sentiments, I believe I have shown how the skeptic can agree to having a duty by using the example of love, leaving open the possibility for other examples to follow.

CHAPTER 5 CONCLUSION

5.1

What I have attempted to show here has been the general framework for the duties of a free person. First and foremost, a free person is one who can choose her own ends, her own goods, for her own reasons. If she is to be judged against some particular standard, it is because she holds *herself* to that standard; she is not *required* to be in any game she does not feel like playing. This is plainly at odds with the truth-possessed tradition of illuminating what is objectively best, with freedom, therefore, being the liberation from ignorance and caprice, and the pursuit of the ideal. The existence of an objective ideal, or even some ordinal ranking, entails that agents in fact are not free to choose their own ends; rather, they are all *stuck* in the same game, the one of seeking the objective best. But this truth-possessed framework is required in order for one to feel a sense of duty. This is because duty, as I have tried to argue, requires one to prioritize something which is beyond one's own preferences and desires. Doing the right thing is not easy, is often unpleasant, and is certainly not self-serving, at least primarily. My task, then, was to try to square this particular circle, to show that a free person is in fact duty-bound. In doing so, I attempted to represent freedom in the strongest possible way, using the skeptic as its flag-bearer. Stick *him* in some game and he will storm away from the table after knocking all the pieces onto the floor – that is, unless it is a game he would freely choose to play himself.

My methodology was therefore to find such a game that appealed to the skeptic, the playing of which he would prioritize over himself. In this, I took a cue from the rationalists, and Korsgaard in particular. Hers was an attempt to ground the norms of agency in its constitutive principles. However, as I believe I have demonstrated, she could not simultaneously *allow* the skeptic to freely choose his game and *entice* him into playing hers. This is because her argument hinged solely on *what it means* to be an agent: depending on how she is interpreted, agency is either too thin a concept to give her the prescriptions she was after, or a concept too thick to convince the skeptic. What she *did* convince him of was what is involved in playing a particular game well, but not that he should play it, which is a crucial omission. He has to be convinced of his own ends, his own values, in order to be convinced of his duty. The rationalist believes that reason alone will lead one to accept one's duty. We saw that reason could convince the skeptic to conform to the correctness conditions for acting for the sake of what he *does* value, since that is what it means to value that thing. But we also saw the difficulty in attempting to extract values from thin concepts such as agency. And more importantly, we saw how the skeptic stood unfazed even by statements of *fact*; why should these affect his dispositions? Reason, he says, allows him to be effective in seeing to his own chosen ends, and he would rather not bother with any others. He is free to do as he wishes. My project, therefore, was to try to find out what he uses all this freedom *for*.

The task then turned inward to the contingencies of the skeptic's constitution. Like Korsgaard, I sought a thick enough characterization of the skeptic's (and everyone's) volitional characteristics in order to entail some duty, but I differed from Korsgaard in two key respects. First, the thickness of the characterization was epistemically justified

on an empirical basis rather than a rationalist one. One of the things the analysis of Korsgaard was meant to show was the futility of trying to *generate* feelings of duty using reason alone – although it could transfer them. At his most stubborn, the skeptic stood steadfast in defiance of even objectively true moral duty (which was only supposed hypothetically, of course). Appealing to empirical truth to ground the skeptic's feeling of duty, there was no way to bridge whatever truth a rationalist was espousing to the feelings of the skeptic. Why should he care? Of course, the same could (and should) be asked about any empirical truth which is purporting to affect the skeptic in that way as well. This leads to the second difference between my empirical account and Korsgaard's rationalist one, which is that the skeptic's duty is grounded not in the *fact* of his constitution, but in the constitution itself. The truth I used to exemplify this idea was Frankfurt's theory that one acts in the interests of that which one loves. Therefore, it is not the *fact* of one's loving something, but the love *itself* that makes one feel a duty. This kind of truth, that we are loving creatures, cannot be garnered from rationality alone but must be discovered about the world. Together, these two differences help to avoid the problems that Korsgaard had. Using empirical facts means that we can have some common basis for discussion, not one that is devoid of practical import, but one from which substantial conclusions can be drawn.

5.2

It should probably be asked, further, just what *kind* of duty has been established here. Korsgaard's was an attempt, following Kant, to ground both moral and rational duty *and*

to do both in ways that would convince the skeptic. Rather than take up Korsgaard's whole project, I have started at the back end, asking what kind of duty the skeptic would accept, generally speaking. Once this is established, we can then see if moral duties, by whichever conception, intersect with those accepted by the skeptic. It is in this way that I believe I have put us in a better position – though there is still much work to be done – to answer the question of whether the skeptic would ever accept a duty that we can reasonably call a moral one. Frankfurt is very clear that “[t]he authority that stands behind the imperatives of love is not at all the same... as the authority with which moral imperatives are imbued,” (OC 170) in a paragraph that more repeats this point than argues for it.²⁸ The point may or may not be true, but what I have tried to show is that the skeptic can only be convinced by something which is deeply felt and integral to his will in the way that his love is – so that his actions are truly autonomous. I have not meant for love to be the only such possibility, but I do take it to be a paradigm case. Also, crucially, I have meant for the ubiquity of love to be such that the skeptic cannot opt out by asserting that he does not love anything; we are all loving creatures.²⁹

Love's key aspect is the normative *grip* that it has on the skeptic. This is because gaining the skeptic's agreement to his duty is what I set out to do. Following Korsgaard, I

²⁸ The full paragraph: “The similarity between the ways in which we experience and respond to the requirements of moral obligation and to those of love tends to support a rather common presumption that these requirements are essentially of a single kind. However, this presumption is incorrect. The authority that stands behind the imperatives of love is not at all the same, in fact, as the authority with which moral imperatives are imbued. Betraying what one loves is frequently taken to be fundamentally a moral offense; but it is not. The commands of love are not moral imperatives. The necessities that grip us in the one sort of case have grounds different from the necessities that grip us in the other.” (OC 170)

²⁹ Here, it is wise to recall Frankfurt's discussion of self-love. So the misanthrope, who claims to love nothing, must also not love himself in order to claim to be exempt from the grips of love. My claim here is that the true misanthrope (who loves nothing, not even himself), while conceptually possible, is not someone who actually exists.

took this to be an integral facet of duty; however, one might, alternatively, assert that the skeptic cannot be so convinced of any duty that really matters. Yes, loving is a volitional necessity for him, but he could love some truly awful things. Ignoring the (very important) question of where his particular love comes from, it could be argued that love, or anything that would similarly grip him, cannot be connected satisfactorily with what we might pre-theoretically think are his duties – moral duties, primarily. His skepticism prevents him from being moved by these truths.

But this could lead to a quite problematic position, which is to hold the skeptic's duty over his head anyway: right is right, whether one feels it or not. Here, we would do well to remember the lessons of the (ongoing) debate between the truth-possessed and truth-pursued views of freedom. While I have not presented the skeptic as having any one value set in particular, I have presented him as someone who can be reasoned with; for example, he was persuaded by the rational requirement of meeting the correctness conditions of whatever he was trying to achieve. And so while he was, following Mill, rightly skeptical of any “final” truths, and especially those which carried prescriptions with them, he was at least open to entertaining them if there were sufficient epistemic justification. If he does *not* agree with a particular duty, then it is because he cannot fully accept the supposed truth upon which it is based; it does not *feel* right to him. (Remember, it was his actual love, and not merely the fact of his love, that made him agree to his duty.) As Mill would argue, the political implications of holding such a duty over him are severe, which makes doing so highly suspect.

The alternative is to develop a “thicker” conception of agency which has empirical justification, which is the project I started in chapter 4 with love. Mere love

may not be enough to ground any “real” duty since, again, one could love some very bad things; but maybe we can dig a bit more, empirically, to see if there are any more duties that grip him. This is where Hume and Smith and the moral sentiments would come into play, if it could be argued that the skeptic is similarly gripped by them. The existentialists might also have a thing or two to teach us about this thicker conception of agency, since, among philosophers, they are probably the most in-touch with the first-person perspective.

In general, the skeptic is difficult to convince, but not impossible, and he is open to new evidence. He is *reasonable*, even if he does not share your values. As such, he is the prototypical Humean foil. I believe I have shown that he can feel some sense of duty which is grounded in an empirical fact which entails it. He is a free person, and vehemently so, but he is not *only* that. Like all of us, he is encased in a human body, with all that entails. If he can feel duty it is because *we* can, inherently. Thus, his duties, the duties of a free person, give us a sense of what binds us, of what gives us a sense of something greater than ourselves and makes us look beyond our own self-interest, and of what matters to us.

5.3

In closing, it is probably worth placing the story of duty that I have outlined into the philosophical tradition more clearly. I have tried to outline the story in a reasonably intuitive way, but it is not unfair for one to demand a more technical explanation – particularly when there are important arguments in the tradition that draw on distinctions

I have largely avoided. Placing my story in this context should therefore clarify just how (and if) those arguments come to bear on it.

Specifically, one might ask about the cognitive versus the conative aspects of the story. As I have outlined it, the skeptic feels his love for this or that and then feels a corresponding duty toward it, with, as I said, no further work being done by any knowledge he may have about his love. This was how I got around the worry of the Humean (sophisticated) skeptic, so it should not be surprising that this fits in nicely with Hume's (and Smith's) idea of the moral sentiments. This is a conative story, one about feeling. But I have also appealed to the truth-possessed tradition of freedom, contrasting rational truths (specifically those about agency) with empirical ones, illustrating how they can allow someone to converge on an action or state of affairs that she finds best. This was to convince the libertarian and naive skeptics, allowing them to choose freely and for their own reasons, lest they be alienated from their actions. Frankfurt's idea of love was then brought in to show that their own reasons were (among other things) the interests of their beloved. This is a cognitive story, one about thinking. The libertarian skeptic believes that he has a duty to his beloved. (Recall that he believed in no such duty toward things he desired.)

If I had presented only one of these then at least one of the skeptic's personalities would have been unsatisfied: the libertarian skeptic would have been unconvinced by the conative story, and the sophisticated skeptic would have been unmoved by the cognitive story (as was Hume). But it might legitimately be asked whether there has been conflation or equivocation on my part of these two aspects of duty, in order to get the desired result of convincing all the skeptics. In fact, this kind of criticism would parallel

my criticism of Korsgaard: where she used whichever conception of action suited her purposes at the moment, I used different conceptions of duty, depending on which skeptic I was trying to convince.

While this would indeed be a delicious irony, I am afraid it is not to be. This is because love – my paradigmatically illustrative example of satisfying *all* the skeptics with their duty – provides *both* the cognitive and conative components of duty. Being a kind of feeling, love gives the sophisticated skeptic the conative component that no mere fact would give him. But of course, love is not *just* a feeling. Love – as opposed to, say, mere infatuation – is evaluatively loaded. To love something, for instance, is to see it as *worthy* in some way, worthy of being prioritized, at least by the person who loves it, and this creeps into the cognitive component of the story. If A loves x, then not only does A feel duty-bound to x, but A also believes that she should act in the interest of x. I believe that (at least) both of these follow from the idea of love. It is not that I have conflated the two senses of duty; rather, I have established a common basis for both of them. Since my goal was to establish the duties of the skeptic – and thus of the free person – they were both needed.

Both the cognitive and conative duty established by love are, unsurprisingly, quite personal. This is appropriate for the conative duty; A might feel duty-bound toward x, but B might feel no such thing, and this is understandable given the differences between A and B. But if A *believes* she has a duty toward x, would she not believe that B should have the same duty, at least if B were in her shoes? What is it that A believes, exactly, when she believes she has a duty toward x? And to whom is it applicable?

These are tricky questions. There are two interpretations one might have of the cognitive component of one's duty toward one's beloved. First, there is something like the following argument: A loves x, and anyone who loves something has a duty toward it, and therefore A has a duty toward x. But this is not quite right because of the evaluative component of love that I mentioned. In loving something, A sees it as an object worthy of her attention, efforts, respect and so on. A might not see another object, y, in this way at all. Now, if someone *else* (say B) loves y, would A believe that B has a duty toward it? (Well, A believes that B *believes* he has a duty toward it – because, as I have tried to show, everyone up to and including the skeptic believes that they have a duty toward their beloved. But this does not answer the question.) And further, if B does not love x and A does, does A believe that B has a duty toward x?

Contrast the above argument with this interpretation of the cognitive component of duty: I love x, at least in part because I see it as an object worthy of my love, and because it is so worthy I believe I have a duty toward it. On this interpretation, the answer to one of the above questions is that A does not believe that she has a duty toward y, since she sees it as unworthy. The other question – of whether A believes that B has a duty toward x – is still tricky. If she *does* believe that B has this duty, then the "worthiness" she believes about x she thinks is something universal – not just aligned with *her* values, but the values of all. As the preceding chapters should suggest, this need not entail that this worthiness is the result of some "objective good" or something of the sort; rather, it could be a result of contingent but universal facts about us, that we do all have certain values. In fact, given the discussion of how to convince the skeptic, if A has any leg to stand on trying to convince B of x's worthiness, then she needs to appeal to

something that is within B already – his values, priorities, and so on. Since the skeptic was, following Mill, wary (to say the least) of the supposed objective valuations of truth-possessed theories, trying to convince him of something's "objective worthiness" or something of the sort is probably not the way to go. (And it is not a far leap, if it is a leap at all, to go from this to a sort of moral objectivism, about which it will suffice for me to say that its contentiousness and implausibility lead me to want no part of it.) On the other hand, if A does *not* believe that B has a duty to A's beloved x, then on some level A understands how B can have a different valuation of x than she does. (Perhaps x is something that only a mother could love, and A is x's mother.)

Given the above discussion, in the context of duties from love, it does *not* seem to follow that in general a duty for one (e.g., A's duty toward x) is a duty for all (e.g., everyone's duty toward x), even though it does follow, in a sense, that everyone in the same situation has the same duty (A's duty toward her beloved, B's duty toward his beloved, and so on). It might be possible that *some* duties are so universalized, but these would require the kind of backing that I argued would be needed to convince the skeptic.

The duty I have outlined is, as I said, quite personal. As such, I have steered away from “objective” values in the sense that I have discussed, though I have left room for universal ones if they could be empirically established. Therefore, I consider my arguments immune to attacks on objective morality or goodness or other things of the sort.³⁰ (For example, there have been no inferences from “I have a duty toward x” to “it is right/good that I serve x.”) And for reasons I mentioned, the position of the skeptic (particularly his skepticism about “final truths”), and my requirement that he be satisfied,

³⁰ I have in mind here J.L. Mackie and his argument from queerness. My great thanks to Greg Scherkoske for prompting me to address it.

in fact steers us *away* from this sort of program. So while there is a cognitive component – a belief one has about one’s duty – it is not far-reaching enough to be problematic in this respect.

BIBLIOGRAPHY

- Dick, D. (2009). *Ethics and the Possibility of Failure: Getting it Right about Getting it Wrong*. PhD Dissertation, University of Michigan, Ann Arbor. Accessed at http://deepblue.lib.umich.edu/bitstream/2027.42/63687/1/ddick_1.pdf.
- Enoch, D. (2006). Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action. *The Philosophical Review*, Vol. 115, No. 2 (Apr., 2006), pp. 169-198.
- Frankfurt, H. (1999). *Necessity, Volition and Love*. Cambridge: Cambridge University Press.
- Hume, D. (1739). *A Treatise of Human Nature*. Accessed at http://www.gutenberg.org/files/4705/4705-h/4705-h.htm#2H_4_0075.
- Kahneman, D. (2011) *Thinking, Fast and Slow*. Doubleday Canada, a division of Random House Canada Limited.
- Korsgaard, C. (1997). The Normativity of Instrumental Reason. In G. Cullity & B. Gaut (eds.), *Ethics and Practical Reason*. Oxford: Clarendon Press.
- Korsgaard, C. (2009). *Self-Constitution: Agency, Identity and Integrity*. Oxford: Oxford University Press.
- Mill, J.S. (1859). *On Liberty*. Indianapolis: Hackett Publishing Company.
- Rosen, G. (Manuscript). *Normativity, Meaning and All That*. Unpublished.
- Rousseau, J.J. (1755). *Discourse on Political Economy*. (G.D.H. Cole, Trans.). Accessed at <http://etext.lib.virginia.edu/etcbin/toccer-new?id=RouPoli.xml&images=images/modeng&data=/texts/english/modeng/parsed&tag=public&part=all>.
- Rousseau, J.J. (1782). *The Social Contract*. (G.D.H. Cole, Trans.). Accessed at

<http://www.constitution.org/jjr/socon.htm>. (Original work published 1762.)

Russell, B. (1946). *History of Western Philosophy*. London: Routledge Classics.

Schaeffer, J., Burch, N., Björnsson, Y., Kishimoto, A., Müller, M., Lake, R., Lu, P., and

Sutphen, S. (2007, September 14). Checkers is Solved. *Science*, Vol. 317 no.

5844, pp. 1518-1522.

Smith, A. (1759). *The Theory of Moral Sentiments*. Toronto: Penguin Books.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell Publishers.

Velleman, J. D. (1992). What Happens When Someone Acts? *Mind*. New series, Vol.

101, No. 403

Watson, G. (1985). Free Agency. *Journal of Philosophy* 72 (April):205-20