THE STRUCTURE OF PRACTICAL RATIONALITY

by

CARL J. HAMMER

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy,

The City University of New York

2009

© 2009

CARL J. HAMMER

All Rights Reserved

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

	Douglas Lackey
Date	Chair of Examining Committee

	Iakovos Vasiliou
Date	Executive Officer

Stefan Bernard Baumrin

Catherine Wilson

Douglas Lackey

Iakovos Vasiliou

Noël Carroll

Supervision Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

THE STRUCTURE OF PRACTICAL RATIONALITY

by

Carl J. Hammer

Advisor: Professor Stefan Bernard Baumrin

Many pressing metaethical problems can be conceived as a need for placing a kind of meaningful and objective morality into an integrated and explanatory worldview, and this requires a constructive explanation of moral obligation. There are two major problems for giving such an explanation. On the one hand, moral obligations must be grounded in a general scheme of practical normativity; otherwise, they can have no authority. On the other hand, moral obligations must arise from social relations; otherwise, they lose their character as demands that a moral community has the authority to enforce.

To explain practical normativity in general, I implement and refine a certain kind of explanatory strategy for normativity, which has been developed by J. David Velleman and Christine Korsgaard – *constitutivism*. To use this strategy, agency and action are conceptually analyzed in terms of a constitutive aim and it is argued that this aim has supreme authority for all who qualify as agents in this technical sense. I argue that a rational agent must aim at systematization of the agent's commitments, and that this aim has the authority to determine correct decision-making for the agent.

To show how this can be worked into a theory of moral obligation with its special social character, I argue first for Stephen Darwall's conception of moral obligations as arising from second-personal accountability relations. Then I argue that having a commitment to participation in the moral community – the social group of individuals who jointly subscribe to mutual accountability – is a plausible condition of human nature (what most people are like). Further, it is also plausible, I argue, that for most people this commitment has an authoritative systematic position within one's scheme of commitments. Moral obligations arise directly from the accountability relations within the moral community, and so the authority of one's commitment to the moral community translates into the authority of moral obligations for that individual.

Acknowledgments

I would like to thank my advisor, Stefan Baumrin, whose support, advice and encouragement were so instrumental in planning and executing this project as well as developing as a philosopher throughout graduate school. Through his insightful lectures, remarks, questions and many forms of antagonism he pushed me to think harder and more profoundly about the issues with which I was grappling. His guidance and support as a mentor have helped me gain my footing in academic philosophy in numerous ways. For all of this, I thank him.

I would like to thank my parents, Don and Linda, for a life of stimulation and encouragement. Their guidance and example has led me to take thought and learning seriously. Their love and understanding gave me a foundation that made everything I have done possible. I hope to emulate them in my own way.

I would like to thank my friends who have helped guide me through this with their insight and discussion and have kept me sane with their quick jokes and sympathy. For this I thank Vic Crome, James Dow, James Snyder, Damien Dupont, Kris Kemtrup, Keota Fields, Mary Thorne and Brian Young.

Lastly, I would like to thank my wife, Jennifer, who has always believed in me. Her love, support and willing ear have helped me through the trying process of graduate school and dissertation life beyond measure. She has shaped me and my life in countless ways for which I owe her the greatest debt of gratitude. My heart is always with her.

Preface

This work is an attempt to explain how it is that there can be and is such a thing as moral obligation. The explanation is conceived as an answer to two primary questions: "How can there be such a thing as practical normativity?" and "How can practical normativity give rise to a system of accountability relations that give a moral community the authority to demand compliance with various social-cooperative norms?"

The first question requires some kind of metanormative constructivist answer that is conceptually independent of and prior to moral relations. I answer it with a form of *constitutivism* – the metanormative explanatory strategy that uses conceptual analysis of agency and action to discern a constitutive aim of agency and then argues that such an aim can form the foundation of a system of practical normativity. This strategy has been developed by J. David Velleman and Christine Korsgaard. Velleman has argued that the constitutive aim of action or agency is self-knowledge (Velleman 1996, 694-726;Velleman 2000), while Korsgaard has argued that the constitutive aim of agency is the unity of the agent (Korsgaard et al. 1996;Korsgaard 2002). My constitutivist theory differs in that I argue that the aim of agency is systematization of the agent's commitments, which is similar to Korsgaard's unity of the agent aim, but is more formal, has clearer implications for how practical thought is to proceed, and further breaks down the various strands involved in an agent's decision. Whereas Korsgaard seems only to argue for the Kantian claims that one must act on a principle and that one must value all of humanity equally, I argue that acting on principles and the role of rational deliberation

require that one have a system of principled commitments and that one orders and restructures them as best as possible in terms of their systematic power. (I do not claim that there is any formal requirement to value of all humanity equally.) In addition, I hope to clarify and improve upon the general constitutivist strategy through a more direct, dedicated and thorough treatment of the strategy itself. To do this, I first discuss in very explicit terms at the outset of my constitutivist arguments just what the constitutivist strategy is and how it is supposed to work. Second, I carry out a rigorous and formal argument for my particular constitutivist explanation of practical normativity. I believe my presentation of, and my arguments for my constitutivist theory of practical normativity is a more detailed and careful execution of the necessary steps of such an explanation than are the arguments either Velleman or Korsgaard have given to date.

The second question refers back to the answer to the first. We want to know how the normative framework being proposed gives rise to a system of social-accountability-based obligations, because these two elements come from different places. A general framework for practical normativity must arise from the first-person deliberative standpoint, whereas, a system of authoritative social demands must arise from social relations and a more socially-oriented standpoint. Here I take after the arguments of Stephen Darwall, who has articulated a conception of moral obligations as essentially certain kinds of demands backed by second-personal social accountability relations. I argue that we should accept this conception of obligation, and I further fill it in by describing relations between moral obligation and the moral community. However, I also argue that Darwall does not successfully answer the question of how a system of such social-accountability-based obligations arises from a general normative framework. This

failing is made inevitable, because he tries to explain everything, including the original source of normative authority for obligations, from the social standpoint.

I answer the second question with a combination of the systematization standard of practical normativity and plausible claims about human nature. If people in general have a deep and authoritative commitment to participating in the moral community, then this will give rise to obligations that spring from the relations of mutual accountability that bind the moral community. So, given the plausible claims that people in general (this need not be universal) have an authoritative commitment to the moral community and that people negotiate relations of accountability that capture social-cooperative norms, we have an answer to the second question. Such facts along with the normative framework would make for authoritative obligations arising from social-accountability relations.

What results is a two-tiered system of right action. There is, in the first place, right action proper, which is just whatever is strictly authoritative for a given individual in the situation at hand. But there is also the separate category of moral obligation, which is whatever the individual is accountable to the moral community for doing, regardless of whether the obligation has full, overriding authority in the situation at hand. What an individual is accountable for depends on the individual subscribing to the overall system of social accountability, and whatever the moral community has constructed in accountability relations. A person may find oneself in situations where one's moral obligations diverge from what is strictly authoritative for that person, as can happen, for example, when social demands come into strong conflict with demands from family or friends, or come into conflict with aspects of life the person holds most valuable.

This system of right action and of moral obligation is part Kantian, and part Humean. It is Kantian in that right action comes from the structure of agency, agency requires acting on principles, and that one's rational deliberations ultimately determine whether acts are right. It is Humean in that the particular actions that are right is entirely contingent, and any general rules we might form about this could only be empirical generalizations based on human nature. This is a kind of *agent-relativism*, meaning that whether an action type is right depends on the agent performing it, even if we include a full description of the context. However, it is not a kind of *judgment-relativism*, according to which things are only right or wrong according to a framework of judgment or a particular perspective. Actions are right or wrong for particular agents in particular situations in an absolute sense – it does not depend on the perspective of the judge. So, according to the theory I defend, the idea that there is no single true morality is correct in the agent-relativism sense, but not in the judgment-relativisim sense.

The Structure of Practical Rationality

Table of Contents

Preface vii

List of Formal Arguments and Proposition Schemes xiii

Introduction 1

Part I: The Problem of Moral Objectivity

Chapter 1: Moral Skepticism and the Internalism-Externalism Dichotomy 16

The problem of moral objectivity is delineated and discussed. It is argued that the greatest problem for moral objectivity lies with the potential for a satisfactory account of practical normativity. Also, certain constraints on giving a satisfactory account of practical normativity are argued for in the context of the reasons-internalism debate.

- **I. Setting out the parameters for an explanation of morality** 16
- **II. Normativity as the core of what is mysterious in morality** 18
- **III. Constructive Explanation** 31
- **IV. Classical realism unsatisfactory** 35
- **V. Obligation by agreement unsatisfactory** 38
- **VI. Constitutive internalism** 43

Part II: A Theory of Practical Normativity

Chapter 2: The Constitutivist Approach to Practical Normativity 55

An initial account of practical normativity is argued for. First, an account of agent-causation – how behavior can be caused by an agent rather than a mere part – is presented. This is coupled with constitutivist arguments – arguments that normativity springs from what is constitutive of agency – that build this up into an account of practical normativity.

- **I. Introduction to Part II** 55
- **II. Constitutivism** 60
- **III. Actions performed by agents** 80
- **IV. Initial presentation of the formal argument for practical normativity** 86
- **V. Agent-Causation** 92
- **VI. Unity Thesis** 112
- **VII. Remaining Premises** 116
- **VIII. Summary** 121

Chapter 3: Systematization and the Standards of Rational Deliberation 124

The general argument from Chapter 2 is refined and expanded. The argument now concludes that actions are right insofar as they result from deliberation that successfully tracks an aim to systematize one's commitments. This conclusion is clarified, at which point a theory of practical normativity that meets all of the constraining criteria from Chapter 1 is arrived at.

I. The goals of this chapter 124
II. The structure of rational deliberation 126
III. The consequences of the argument for systematization 156
IV. Resolution of the problem of practical normativity 178

<u>Part III: Social Accountability and Moral Obligation</u>

Chapter 4: Moral Obligation: Normativity and Social Accountability 183

To move from the explanation of practical normativity proposed in Part II to an explanation of moral obligation, Darwall's theory of second-person accountability is analyzed. The issue is presented as a puzzle: how to reconcile the various arguments and goals already presented and Darwall's arguments that moral obligation must be second-personal. This is further developed by arguing that social accountability actually does occur and that the system is, therefore, a complete system of moral obligation.

I. Introduction to Part III 183
II. Morality as a social phenomenon 186
III. Darwall's arguments 193
IV. The shortcomings of this view 210
V. The puzzle 219
VI. The solution and resolution of the two conceptions of morality 226

Chapter 5: Contingency, Subjectivism and the Moral Community 236

It is noted that the theory of moral obligation which has been argued for leaves obligations contingent and merely general. The charge that this and other characteristics of the view leave morality subjective is met by distinguishing relevant senses of 'subjective' and rejecting them in turn. The idea of social accountability is developed in terms of participation in a moral community. Finally, the overall picture of morality being advocated is sketched and discussed.

I. A complex picture of contingent practical life 236
II. Substantive Subjectivism 240
III. Formal Subjectivism 247
IV. The final picture 257

Bibliography 266

List of Formal Arguments and Proposition Schemes

Argument 1 72
Enoch's First Argument

Argument 2 72
Enoch's Second Argument

Argument 3 73
Constitutivist Argument that Practical Normativity only Applies to Agents

Argument 4 78
Argument for the Explanatory Power of Constitutivism

Argument 5 91
Unity Version of the Practical Normativity Argument

Argument 6 111
Agent-Causation Argument

Argument 7 128
Systematization Version of the Practical Normativity Argument

Argument 8 216
Argument that the Supremacy Thesis Must Come from a General First Person Standpoint

Argument 9 222
Social versus Deliberative Puzzle of Moral Obligation

Introduction

How is morality possible? That there are objective and authoritative standards for what can be demanded of a person's conduct is mysterious from the perspective of theorizing about a comprehensive worldview. While it seems very real, and, in practical terms, nearly impossible to reject the assumption or ideal of objective practical standards that our moral reflections track, it is utterly mysterious where such standards might derive from that would give them their force or authority. Furthermore, if there are moral elements built into the structure of the world in some sense, it is still baffling how we could have knowledge of them, and how they could have any meaning for our deliberations about what to do.

In this essay, I will attempt to give a certain kind of answer to the question, "How is morality possible?" On the one hand, morality, as genuine, objective normativity which bears on important aspects of our lives, seems very real. It seems as though there really are things which I should do and things which I should not do in some unqualified sense. Similarly, it seems as though there really are things that other people should and should not do. Furthermore, it seems as if there are things we should do that we owe to each other. The compelling nature of this appearance is perhaps the driving force behind the various forms of moral realism one finds.

When I consider actions which are typically taken as cruel, horrible and unjustified, like human trafficking, murder, or simply making gratuitous insulting remarks to a stranger, it seems that these actions are simply wrong: they should not be done and agents owe it to others not to perform them, such that it is right for the moral community to enforce rules against such behavior. This appearance is entirely compelling, even after the ingestion of a healthy dose of skeptical philosophical argumentation. In addition, I cannot help but blame others and myself for acts that seem to violate what I and others owe to each other. I also demand, through my attitudes and actions that others and I act in accordance with various norms I take us to be socially accountable for. More importantly, from the first-person perspective of an agent who is choosing what to do, thinking of some actions as right and others as wrong is unavoidable. In deliberation one has to assume that there are right and wrong decisions. This is what one is trying to figure out through deliberation. From the practical standpoint of deliberation, I cannot entirely believe that there is no value one way or the other to what I do. To think in such a manner is to give up one's agency, at least with respect to the current decision.

So far, these considerations bear only on how I have to think, and not on whether this thinking is correct in its representation of the world. However, the appearance of correctness here is very strong. Even for those philosophical skeptics of morality, the need to worry about how other people are being treated and what direction society is going in, and so on, considered in terms of right and wrong, is often too appealing to deny. It appears that there really are genuinely objective moral standards.

On the other hand, it is a mystery where such normativity which has bearing on our practical lives should derive its legitimacy. Contemporary philosophy is dominated by acceptance of the modern scientific worldview, which seeks to put empirical science in a leading role in the investigation of the world of our experience. Not only that, but it seeks a kind of metaphysical parsimony and, most importantly, a thoroughly explanatory approach to philosophy, wherein unexplainable phenomena are serious problems for a theory. I take the core naturalist idea to be that we should seek a thoroughly integrated explanatory account of the world of our experience, upon which we place no insurmountable barriers to explanatory connections.

To take what is generally considered a reasonable view of the world and the nature of the human being, it is clear that, at least at bottom, humans are composed of scientifically describable material. Furthermore, it seems that all of our behavior and behavioral tendencies are, at least in principle, explainable using only scientific (non-evaluative) descriptions (Harman 1977, 3-10). This seems to leave no room for morality to get in. Morality, as a kind of governance of right and wrong action, cannot be explained anthropologically or psychologically, for this is merely to explain it away. To provide a purely anthropological or psychological explanation of morality could only be to explain moral practice or its origins. While this may be an interesting and important task, this cannot capture the normative element of morality, and so cannot be an explanation of morality as a kind of governance of right and wrong action.

What could there be that makes it true that we have various authoritative moral obligations? If we consider what features of our psychology and what features of the world around us would provide the basis for moral obligation, it is difficult, to say the

least, to come up with a satisfying answer. Our decision making process seems only to create choices and the *assumption* of right and wrong. The world apart from our agency seems only to provide data to consider in making choices, but nothing which provides a schema for right and wrong. How is morality possible?

In my view, this problem is exactly analogous to the problem of the synthetic *a priori*. On the one hand, there is an extremely compelling appearance of the reality of synthetic *a priori* knowledge. On the other hand, it is a great mystery how there could be such knowledge. The mysterious nature of this knowledge is what has led many philosophers to reject its legitimacy. Similarly, the mysterious nature of moral obligation has led many philosophers to reject the legitimacy of *it*. I will try to explain moral obligation in terms which are less mysterious, and thereby try to answer the question of the possibility of morality, just as many others have already tried to do. Perhaps the view I will present will be seen as an unsatisfactory answer, because it does not establish morality in all of its potential or perceived glory. What I will say may strike some as disappointingly subjectivist, but I will argue that it is not subjectivist in any way that makes it inadequate to capture the basics of our idea of morality and moral obligation and its authority.

To explain moral obligation there are two major challenges, as I see things. The first challenge is to explain practical normativity in general. This must involve using practical concepts and human nature to explain how there can be and are standards of right and wrong action that are authoritative in an absolute sense, about which an agent can be wrong, and which can play a meaningful role in how an agent deliberates about what is right and what to do. These criteria ensure that the explanation integrates

practical normativity within a larger, explanation-friendly view of the world without being reductive in a deflationary sense.

The second challenge is to explain how practical normativity gives rise to a system of accountability relations that give a moral community the authority to demand compliance with various social norms. The sense of 'morality' that we want to explain consists of moral obligations that arise from accountability relations in that one who has an obligation is responsible to someone or other (possibly to oneself). Further, that these relations bear a connection to the moral community such that an obligation always implies the community's authority to enforce the obligation, even if only through blame or resentment. Without this social accountability aspect obligations lose their character as being rightfully enforceable in general. It is important that the explanation of social accountability relations occurs within the normative framework that has, presumably, already been explained, because without a general normative framework, we cannot explain the authority these relations are supposed to entail.

So, the way these two challenges are met will have to be combined into one continuous strategy. This will require the joining of two different perspectives. We need an explanation of practical normativity for the deliberative standpoint, and we need an explanation of the normativity of social accountability for the social standpoint. From one perspective, we must work to build an agent-centered account of practical normativity, because only by doing this can we arrive at a truly authoritative system of practical rationality and right action. From a different perspective, we must try to capture what it is about our social relations that make certain things social requirements in a way that has a moral character.

To meet the first challenge I analyze the conceptual structure of agency, starting from the basic notion that at the most fundamental level agency is the process of making a unified decision from one's diverse motivational field. This is what constitutes the *agent's* decision. So, explaining how this can happen is to explain agency and account for the ownership of action. I argue that this happens just in case one makes a decision guided by an aim to systematize one's commitments, where *commitments* are explicit and implicit principles of action. These and related arguments concerning action and agency I combine with arguments that purport to show that an aim of agency, such as the aim to systematize one's commitments, can provide a system of practical normativity. The result is a theory of practical normativity that achieves the kind of generality and authority we are looking for.

To meet the second challenge and to combine the social and individualistic perspectives I first argue for Stephen Darwall's second-person accountability conception of moral obligation. This differs mainly in the level of detail from what I have already argued above concerning social accountability, but it also makes explicit a theoretical commitment to obligations arising from within the second-person standpoint, which creates a special case of the tension between the individualistic and social perspectives. I set out the tension between the second-person and first-person standpoints as a puzzle and point out some of its difficulties. The solution I provide relies on claims about human social nature and their integration into the normative framework I argue for to meet the first challenge. This leads to two distinct domains of "right" action. There is, first of all, right action proper, which is just whatever follows from the general normative framework for a particular individual or group, which is not necessarily the same for

everyone. Right action in this sense is not universal, although it is completely objective in the sense that it has true authority in its application. If it is right for me to take the last cookie as implied by my deliberative aim to systematize my commitments and my actual set of commitments, then it is right for me in an absolute and authoritative sense, without restriction to the perspective of any judge or observer. However, there is also a qualified sense of 'right action' that has the qualities of moral obligation. This sense is universal within a moral community – a social group that is mutually committed to joint, universal accountability. Accountability relations from this standpoint give rise to moral obligations that get their ultimate authority from the authority of the commitment of each individual to participating in the moral community. The moral domain must come from the general normative framework, but it gets distinguished by its particular character. The two resulting domains of "right" action each have their place in practical life. The individualistic, absolute domain is what is appropriate to consider on an individual basis, while the social, moral domain is what is appropriate to consider as a social group (although these lines do not hold for the considering and giving of advice for someone else or for some group).

In giving my theory of morality, I address and provide a solution to a number of problems in moral philosophy. First of all, I explain practical normativity. And, as the primary goal of this project, I use this account of practical normativity and other insights to explain moral obligation. In doing so, I also explain what truth there is to the supremacy of the normativity of moral obligation. It is held by many that having a moral duty implies that one has overriding, or conclusive reason to perform the duty. Why this should be so, however, is very hard to account for. One could simply define obligations

as requiring conclusive reasons, but then, without arguments that what should count as obligations actually are accompanied by conclusive reasons, this may simply restrict what counts as obligation. One may also claim that what we already consider moral obligations just are overriding, but without an explanation for why this should be, this claim has little bite. I do not explain a sense in which everything we might like to call a "moral obligation" has conclusive reasons for all rational agents. I do, however, give an account of moral obligations that makes sense of where their normative force comes from and why they should be expected to enjoy normative supremacy.

I also give a sketch of an account of how an agent can cause, control, or own an act. It is a standing problem in the philosophy of action how an agent can perform an action, as opposed to some part of a person causing the behavior. If someone behaves in such a way as to drop out of college, we want to at least leave room for a sense in which that person, the agent, dropped out of school, rather than this always reducing to some desire, impulse or mechanism, which cannot be equated with the agent, causing the behavior. I lay out a conception of agency from which a theory of the ownership of action can be constructed.

As an interesting and difficult combination, without invoking any kind of attitude-independent classical realism, I make room for the changing commitments of an individual and of a society, while also making moral progress clear and explainable. We do not want to rely on a theory of practical reason according to which one's commitments cannot be rationally changed. We want to be able to account, not just for it being permissible to change one's commitments, but for one having normative reasons for doing so. Say Fred is a loner whose top commitment is survival in a war-torn land. This

commitment may form the basis of most of his reasons for action. But now, if he meets someone with whom he could live and start a family, we want to say that it is not necessarily irrational for him to do so, even though it might hamper his aim to survive in a dangerous place. Further, it should, at least in principle, be possible for Fred to have normative reasons to change and reorganize his life in a dramatic way.

Giving such an account, however, can easily fall into tension with the idea of moral progress. If one can change even some of one's most basic commitments rationally, then it is hard to see how individuals or groups could make progress morally or do better than they did before. Something needs to stay fixed according which progress can be measured. Without any kind of attitude-independent realism, it would seem that doing better would have to mean sticking to one's commitments more faithfully. While moral progress, on my system, may not be as linear and straightforward as a classical realist would have it, I can easily explain what it is using the elements of my view. Since practical normativity boils down to what is required for an agent to systematize the agent's commitments, and moral obligation comes to the normativity of a community's shared relations of accountability (roughly speaking), moral progress is the improvement in systematicity of a moral community's shared relations of accountability. This will become clearer after I have discussed these conceptions of practical normativity and moral obligation at length.

Along the way to solving all of these problems in moral philosophy and giving an explanation of moral obligation, my theory takes on a character that has a range of interesting elements and consequences. Some of these features make my theory rather unique. Most striking perhaps is the fact that I separate two important senses of the

objectivity of morality. One sense in which morality may be objective is that the force of the normative character of morality is real and authoritative – that a person really does have some particular obligation, say. The other sense in which morality may be objective is that for any obligation, every rational agent would have that same obligation in the given situation. These two senses can come apart without the first sense being at all diminished. My theory shows this to be true, because according to my theory moral obligation is real in that some people (most) really do have moral obligations in a way which is not less forceful than on anyone else's view, but it is also the case that these obligations are contingent. The obligations are contingent, because it is not necessary that any arbitrary agent would have the given obligation in the given situation. Although, presumably, I have an obligation not to light a random stray cat on fire or to torture people for fun, this is a contingent fact about me, and not a necessary fact about every rational agent. Further, moral obligations are merely general in the sense that, while we can perfectly well generalize about what obligations people in our society, or the whole world, have, obligations are not absolutely universal. There will be some rational agents who fail to have various obligations that most of us have. This, of course, is a rather curious feature for a theory that purports to explain how morality is possible, and so I address various challenges to the adequacy of the view pertaining to this.

Another interesting feature of my theory is that, even though I explain practical normativity in terms of the structure of agency and moral obligation in terms of this account of practical normativity, I incorporate a contractualist picture of where the content of moral obligations derives from. Using a model of social negotiation and agreement to capture the content of moral obligations can be taken as a way of trying to

avoid explaining the force of practical normativity. Of course, I reject such an approach, but I do think that contractualism has important insights into how social accountability relations are determined, which in turn form the basis of moral obligations. I believe that what we are accountable for in this moral sense is ultimately a matter of social agreement. The agreement comes, however, not from a simple sort of fixed social contract that we consent to, implicitly or hypothetically, but rather through a dynamic process of negotiation through our attitudes, actions and, most importantly, through our moral discourse. It is the public discussion of morals, on my view, that determines the content of our accountability, although this must be played off of a real basis of our society's commitments that we are jointly attempting to systematize.

Finally, I do all of this without appealing to any mysterious normative or moral properties. I appeal to intentional attitudes and mental processes in my account, but these I take to be far less controversial and mysterious than moral properties or primitive practical normativity. I also do not rely simply on arguing for a cognitivist account of moral judgment, and then assume that if moral judgments can be true in terms of semantic form then some of them surely are.

The overall character of my theory is both Kantian and Humean. While the normative output of the theory may be radically different from Kant's official moral theory, I follow the Kantian approach to moral obligation in that I attempt to explain normativity and the source of normative and moral content using an analysis of the structure of rational agency. The specific form that this will take makes use of a philosophy of action analysis of the constitutive aim of agency – a strategy I call "constitutivism," which has been developed by J. David Velleman (Velleman 1989;

Velleman 2000b; Velleman 2004, 225-238), Christine Korsgaard (Korsgaard and others 1996; Korsgaard 2002; Korsgaard 1999, 1-29), and others (Rosati 2003, 490-527; Railton 1997). This strategy involves arguments that a certain aim is constitutive of rational agency, and then argues that this role in rational agency puts it in position to ground normativity. An analogous strategy is also available for theoretical normativity (Velleman 2000a), but that will not be part of my theory of moral obligation.

My theory is also Humean in that right action, in both the general and the moral sense, is contingent and not universal among rational agents. On the one hand, I argue that the general system of practical normativity based on systematization of the agent's commitments is universal among rational agents – it is constitutive of rational agency, on my view, that one be subject to the normativity of systematization of commitments. On the other hand, however, the content of what is right for a given agent and what moral obligations an agent has entirely depend on the agent's contingent, empirical nature as an individual. If one deliberatively aims to systematize one's commitments, but has no social nature at all, then one would be a rational agent who has no moral obligations, and, presumably, what would be right for that person would be far more self-centered and egotistical than what it would be for any normal adult human.

I set out and argue for this theory of moral obligation in three major parts. In the remainder of Part I, my goal is to set the parameters of the intended explanation of moral obligation. I set out distinctions in how the objectivity of morality may be conceived and point out how the reality of the force of practical normativity is not equivalent to, and does not imply that obligations will be the same for every rational agent in a given situation. I also argue that the real philosophical mystery of morality is where the force

of practical normativity comes from, and how some form of this could supply any content to what a person should do. That is, the mystery of how morality can fit in with a sensible theoretical, explanatory worldview does not primarily concern where in particular the content of obligations comes from; but rather, it is the question of how the force of obligation can connect to meaningful content at all. In Chapter 1, I also define the notion of 'explanation' that is needed to answer the metaethical question, "How is morality possible?" From there, I argue that to give such an explanation requires appealing to the motivation of the acting agent, which of course rules out a variety of objectivist metaethical views as potential answers to the question.

In Part II, my task is to give the desired explanation of practical normativity. In Chapter 2, I begin this undertaking by setting out my overall explanatory argument and supporting its premises. The first step in doing so is to make out and clarify the constitutivist strategy and show how it can work. Thereafter, I argue for various requirements for what it is for an agent to perform an action, and then I sketch an account of agent-causation, or the ownership of action. I also put forward a preliminary argument that the constitutive aim of agency is the unification of the agent that serves as an approximation of the systematization arguments that come later. I follow up these arguments with arguments supporting the other premises in my overall explanatory argument for practical normativity. These premises are largely Kantian in nature, and they put the focus of normativity on the deliberation of the acting agent.

In Chapter 3, I argue for the key premises of my argument, which fit into the overall explanatory argument for practical normativity and replace the preliminary, unity argument from Chapter 2. I argue that rational deliberation is structured around an aim at

the systematization of one's commitments. It takes considerable toil to get to this conclusion, but in the process I make clearer what this means and how it works. Thereafter, I describe the picture of rational deliberation that is needed and explain how it fits with our intuitions about the deliberative process. To help make my view clearer, and to put it in terms of a broader perspective, I compare it and discuss connections to some related views in the literature.

In Part III, I will take the theory of practical normativity from Part II, and I will use it to construct a model of moral obligation. This primarily means addressing the second major challenge to explaining moral obligation that I pointed to – explaining social accountability within the normative framework. I will begin this process in Chapter 4 by looking at Stephen Darwall's account of moral obligation (Darwall 2006; Darwall 2007a, 111-132) and arguing for his conception of obligation as "second-personal accountability." This sets up a problem, because the second-personal, social account of obligation seems to be in rigid conflict with the purely first-personal, rational deliberative account of practical normativity I give in Part II. However, I will argue that the two can in fact be combined and I construct such a hybrid account.

In Chapter 5, the final chapter, I look at the overall picture of moral obligation that I have put together. I consider in detail the various objections that may arise from the appearance of subjectivism that my view may present. I argue that my view is not subjectivist in any damaging way, and that it does constitute an adequate, though minimal system of morality. As a key element of this discussion, I consider the way that public moral discourse is incorporated into the model of moral obligation I am advocating and I

argue that it plays a foundational role. Public moral discourse, on my view, is where social accountability relations and the content of moral obligations come from.

Chapter 1
Moral Skepticism and the Internalism-Externalism Dichotomy

I. Setting out the parameters for an explanation of morality

In this chapter, I argue for a particular conception of what is needed to answer the question set before us, "How is morality possible?" As it stands, it is not immediately clear what kind of answer is really needed. Further, it is not clear what constraints there might be on an acceptable answer. Thus, the first major task in the project of answering our question is to set out a clearer notion of what kind of answer we are looking for and a set of constraints on ways the question might be answered.

I will begin by looking at the target notion of objectivity and look at some possible notions of 'objective morality' that might be what we are looking for. I will argue that one of these in particular lies at the heart of the matter and is the one that both seems most compelling and most confounding. This, to put it very briefly, is the normative aspect of morality. We may ask how morality, considered as a set of rules, could be normative for an arbitrary agent. This amounts to the question of how obligations can have the kind of force or authority, independent of the agent's explicit ends and desires, that seems to be characteristic of what we take obligation to be. Although there are other important features of morality to be explained, such as how the particular content – the particular things that ought to be done – arises and where social accountability comes from, the normativity of morality stands as that in greatest need of

explanation. This is the most important aspect of the objectivity of morality, philosophically speaking, because this is what is most philosophically questionable with respect to the larger project of putting together an integrated explanatory worldview. The normativity of morality, or just practical normativity is what seems the most difficult to integrate explanatorily into a philosophical system that aims to be complete or final.

So, in this chapter it is my task to focus on making out the sense in which we can separate normativity from other aspects of moral obligation and to analyze the importance of normativity and some important constraints on explaining it. In Part II, it is my task to provide the explanation of practical normativity with the constraints outlined in this chapter. In Part III, with an explanation of practical normativity in hand, it is my task to work toward a fuller explanation of moral obligation using the account of practical normativity. This fuller explanation of moral obligation includes other features of morality like the source of its content and its inherent structure of social accountability. Both of these features are essential to moral obligation, although I argue that, by themselves, they are not as philosophically mysterious as practical normativity. They are mysterious however, in how they are combined with an explanatory theory of normativity. Explaining social accountability in terms of a given, independently explained scheme of practical normativity is a weighty task unto itself.

Since an explanation of morality is what is needed, and there are many kinds of explanation, I explicate the notion of 'explanation' that is needed to answer our question in Section III. I define a notion of "constructive explanation" as a variety of conceptual explanation and argue that it is what we are after. At various points I draw on this notion to argue for claims about what does and does not work in giving such an explanation.

Thereafter, I consider two standard views about morality that assign morality a kind of objectivity: classical realism and morals by agreement. Classical realism has the simplest answer for what obligation is – it is a simple *sui generis* property of actions. Actions simply are the subject of binding obligations and that is part of the nature of things. I reject this approach for its refusal to provide any real explanation of normativity. The agreement view can explain many features of morality, but it too fails to provide any real explanation of normativity, as I argue.

In the final section, I look at the role that motivation must play in an explanation of normativity and obligation. This is roughly a discussion of what some have called "existence internalism" (Darwall 1983) or "reasons internalism" (Darwall 1992; Audi 1997). I argue that a version of internalism is required to give an adequate answer to our question and to provide an explanation of moral obligation in the relevant sense. I also argue that, while there is a close tie between classical realism and the denial of strong forms of internalism, the motivation behind these views is greatly dissipated once we see the possibility of an internalist explanation of obligation.

II. Normativity as the core of what is mysterious in morality

If we say merely that we want to explain the objectivity of morality or objective morality, this could include a diverse array of issues. There are many different ways in which morality might or might not be objective and many senses of 'objective' which might apply to morality. Some of these issues of objectivity may be central to the philosophy of morality and some may be of no importance or only of passing interest to those dealing with some of the other issues of objectivity. To clear this up and achieve

greater focus for the project, I will distinguish and discuss three notions of the objectivity of morality, and make clear the extent to which I am aiming at establishing each of them.

When speaking of morality in such abstract and general terms, we can distinguish between two aspects of it: the normativity and the content. The *normativity* is the bindingness, practical necessity, or authority; whereas, the *content* is the substantive matter which applies to particular cases and relates to the empirical. The normativity is the aspect of morality which is directive and makes it correct that the agent choose in accordance with the content, while the content relates empirical conditions to normative categories. The content is that which should be done. The normativity is the force of the 'should'.

Objectivity can apply both to the normativity and to the content. For the normativity to be objective is just for the obligation to be authoritative – for the normativity to really apply. In other words, what is at issue is the *reality of obligation* or the *reality of practical normativity*. For practical normativity to be real is for there to be some force behind what I (or others) *should* do, which is independent of the mere fact that I choose to do it. I add this last clause, because there may be a rather vacuous sense in which I should do something just because it is what I choose to do, where there is no further unpacking to be done of what I choose to do. Analogously, there may be a vacuous sense in which someone else should do something just because it is what that person chooses to do. The reality of practical normativity is, of course, not meant as something vacuous in this sense.

Unfortunately, later on, I must distinguish between obligation and practical normativity in general. In Chapter 4, it is my task to discuss and implement Stephen

Darwall's conception of obligation as practical normativity via accountability (Darwall 2006; Darwall 2007a, 111-132). According to this idea, one has an obligation if one has normative reasons *because* one is accountable to others. For now, I ignore this distinction since it only complicates the point I am trying to make about the overall importance of normativity for morality.

There are at least two senses in which objectivity can apply to content. Perhaps the more basic sense is to have determinate particular content, to which I give the name 'content specificity'. There is objective moral content, in this sense, just in case there are particular actions that some person should or should not do. If there is no particular thing which it is right or wrong to do, then there is no *specific objective content*. If someone should not punch the person in front of them in some particular case, then *content specificity* holds. If, on the other hand, there is *never* any particular thing which anyone should or should not do, then content specificity does not hold.

It is, of course, natural to think that if there is one case in which there is some one thing which a person should or should not do, then there are many cases in which there is something which a person should do and many things a person should not do. In other words, it is natural to think that if content specificity holds at all, then it will hold in many cases, for many people with respect to many actions. If it is true that in some particular case Bob should not punch the person in front of him, then it would be amazing if there were no other cases in which there were any particular actions a person should or should not perform. Of course, it does not *follow* from the fact that content specificity holds in one particular case, for one person with respect to one action that it will hold anywhere else, but if there were only one such singular application of content specificity that would

be very strange and would defy most of what we think about ethics, such as the notions that ethics stems from rationality or human nature or animal nature. These are traits which apply to many individuals and would seem to make the application of content specificity a general phenomenon, if it is something that applies at all. We also might think that punching someone would be wrong because, for example, it is wrong to cause others undue harm in certain cases, and that there are many such cases. This makes it very implausible that there are not other cases in which inflicting harm on someone would be wrong.

Content specificity seems to be necessary for the reality of obligation, since it cannot be that I should without there being anything that I should or should not do. One might think that content specificity is not necessary since it might be that I merely should perform an action from among a certain group of actions rather than one specific action. However, this still implies content specificity since any action not in that group is one which I should not perform. If my obligation only demands that I help society flourish, this still rules out the long-term action of being at war with society in the Hobbesian sense. This argument may be a little controversial, since it depends on one's conception of an action (which I do not intent to discuss in this chapter), but it does not affect what I say about the reality of practical normativity and the varieties of the objectivity of moral content.

The second sense in which objectivity can apply to content, which presupposes content specificity, is universality. *Content universality* is uniform application of moral duties for all rational beings, or all moral agents. This is to say that content universality is true of morality just in case every rational being (however that is determined) should

do the same thing in a particular situation or context, for all contexts in which someone should do something. So, the principle of content universality would have it that if any particular person should pay for his or her sandwich at a deli under normal circumstances, then every rational agent should also pay for his or her sandwich if under the same circumstances. Content universality, understood as a set of rules which determine what should be done by an agent, is equivalent to a function *from* contexts, in which rational agents may be embedded, *to* actions which should be performed by the embedded agent. That is, agents can be situated within contexts which would have particular right (or wrong) actions assigned to them (by the function), if there is a right action for any agent in the context. According to content universality, there is no context where some, but not all agents should perform X. If all situations in which there is a specific thing that a person should or should not do are situations in which any moral agent should or should not do the same thing in that situation, then content universality holds. On the other hand, if there is some situation S in which some person P should do X but there is another person Q for whom it is not true that Q should do X in S, then content universality does not hold. For content universality to hold, duties must be uniform with respect to all moral agents. It must be true, for content universality to be true of morality, that whenever I should not tell a lie, if you were in the same situation you should not lie either; whenever I should help others, you should help them too if you were in my place.

There may be some difficulty in separating the context from the agent in the context. In some cases, we would readily agree that the individual in the role of agent affected the nature of the situation, and thereby affected what were to be done. For

example, there may be times when I should give up my seat on the subway for someone who has a much greater need, and I am relatively able-bodied. There may be many perfectly rational people who would need the seat enough themselves that it would not be right for them to give up their seat on the subway for the person who I should allow to sit. This seems to be a case where the agent and the context cannot be clearly separated. However, we can distinguish the context from the person in the context, by saying that the context includes all facts about the consequences, or perhaps expectable consequences, of the decision while the qualities of the agent include everything that determines how that person makes decisions. So for me to be in a situation in which I should not lie, the question of whether you should do the same thing in that situation comes down to whether you should choose in the same way assuming that you bring your decision making tools into the position, but leave the consequences that lying would have on me in place. For a question of lying, this will generally be insignificant, because what lying does to me when I lie to someone is often insignificant compared to what it does to others or perhaps compared to the inherent rational problems with lying itself. It is usually not me that gets hurt from it, and to the extent that it is, it would have the same effect on you to do it as well. The subway case can be stated more easily. According to content universality, if you are as able-bodied as I am and have the same surroundings, and thus the consequences of giving up your seat were the same for you as they are for me, then you should give up your seat just in case I should.

Traditional and contemporary "objectivist" views seem to either accept all three or reject all three forms of objectivity. Hobbes, Kant, Bentham and Mill all were believers in content specificity and universality. Unless Mill was actually trying to avoid

appeal to obligation with his rather radical empiricism, all of these figures and many others were also realists about obligation. In contemporary philosophy, Peter Railton (Railton 1986, 163-207), Richard Boyd (Boyd 1988) and Russ Shafer-Landau (Shafer-Landau 2006) accept all three forms of moral objectivity, while Gilbert Harman (Harman 1977), J.L. Mackie (Mackie 1977) and Richard Joyce (Joyce 2002) reject all three forms of moral objectivity. In Part II, I argue for a theory of practical normativity that upholds content specificity but not universality. This carries through to the account of moral obligation I develop in Part III. In Chapter 5, I return to the general issue of separating the reality of obligation and practical normativity from content specificity and content universality.

Moving back to the broader distinction between the reality of obligation and the objectivity of content, we can see that this distinction corresponds roughly to the difference in interest between cultural or individual relativism, which is widely discussed in non-philosophical circles as well as in philosophy, and metaethical subjectivism which is more commonly the focus of debate among philosophers. In public discussion, the primary source for moral skepticism is skepticism about moral content, and this concern over moral content often arises from considerations related to cultural and social diversity. One wonders how there could be any one set of rules which is correct when it seems like there are many diverse systems of rules governing social interaction, all of which, in a sense, work. People may talk about how infanticide may be considered acceptable in some cultures but unacceptable in others, or how gender roles can differ greatly from culture to culture. What it is for such a system of rules to "work" may be initially unclear, but this line of reasoning and other, related lines of reasoning are often

employed to argue for a kind of subjectivity or relativism of morality. This is an attack on content universality and sometimes content specificity.

A typical relativism argument is aimed at showing that a given action in a particular context could be right if performed by someone from one culture, but wrong if performed by someone from some other culture. This is contrary to content universality and is what we may call *agent-relativism*, because the supposed rightness is relative to the acting agent. A more serious kind of relativism is *judgment-relativism*, which is contrary not just to content universality, but also to content specificity, because it is the view that all values and norms are relative to a judging party. On this view something that is "right for me" (as a judge) could be "wrong for you" (as a judge).

This is sometimes backed by a relativism-by-diversity strategy that is aimed at arguing that it is the culture or background of the evaluating party (rather than the acting party) that determines any sense of the rightness of action. A good example would be something like the following: It might seem impossible or just implausible that there could be any sense in which one action is better than another without deriving that superiority from the evaluative attitudes of the judging party. One might think that there just is no possible standard for one action being right and another being wrong, without this stemming from an emotionally-based, non-rational moral sense. In short, one might take a sentimentalist as opposed to a rationalist view of morality. Diversity, in this case, is appealed to to show that the kind of attitudes which form the base of moral standards do not appear in the world with any meaningful uniformity. But in the absence of any rational, intuitional, or abstract standard for conduct, the only possible standard would be coincidence of all judging parties in moral verdict. Diversity shows that this does not

obtain and so there is no fact of the matter as to what one should or should not do in a particular case. So, for example, since rightness must be tied to what people agree to, on this line of thinking, infanticide is wrong under the evaluation of a typical modern Westerner, but acceptable under the evaluation of a typical pre-modern Inuit. So, whereas an agent-relativist take would be that killing a baby would be wrong for the modern Westerner *to do*, but acceptable for a pre-modern Inuit *to do*, a judgment-relativist take would be that, whoever did the killing, it would be right *under the evaluation* of one judging party and wrong *under the evaluation* of the other judging party. On the judgment-relativist version, there would not be a particular thing that should or should not be done in the situation, only some things according to some judges and other things according to other judges. So, judgment-relativism is contrary not just to content universality, but is contrary to content specificity as well, which makes it a thorough-going relativism that is much more serious than agent-relativism. It is important for my purposes that agent-relativism not be confused with judgment-relativism, since I ultimately argue for a form of agent-relativism. I should also note, however, that the form of agent-relativism my theory of moral obligation takes is tempered by various social considerations that make it significantly less relativistic than other possible agent-relativisms.

Although judgment-relativism is contrary to content specificity and so also to the reality of practical normativity and of obligation, it can be refuted by direct explanatory arguments for practical normativity. Since giving just such direct explanatory arguments is what I do in Part II, judgment relativism is not a threat to my theory. The kind of relativism arguments considered above are based on a skepticism of content specificity

and normative authority. If, however, agent-relativism is incompatible with content specificity – if, that is, content universality cannot be separated from content specificity – then my theory is in trouble, since I aim to establish a system of practical normativity that is agent-relative. Now, one might think, and one who takes morality to spring straight out of reason would think, that there cannot be particular things anyone should do if they are not things that every rational being in the same position would be obligated to do. This is to say that universality is generally taken as a necessary condition for specificity, and also therefore, normativity. This kind of reasoning, however, is really only a failure of imagination. Just because one has not thought of a way that there could be content specificity and real, authoritative obligation, without content universality, does not mean that this combination cannot be. There is no logical or conceptual reason why content universality would be necessary for content specificity – there is only a failure to imagine what an obligation could amount to other than a relationship between some laws of rationality and particular details of a context of action. If that were the case, if obligations could only spring from rational laws and contexts, then content universality would be a necessary part of any authoritative obligation. The individual agent could not matter, under those conditions, because what is characteristic of the individual is left out of the equation that determines obligations – an equation between abstract rationality and the context. If, however, as I argue in Part II, normative authority and the rightness of actions comes from a relationship between some abstract principles of practical rationality, the context, *and* the determining features of the individual agent, then content universality is not a necessary feature of such authoritative norms. For example, the rightness of an action might depend on the agent's motivations and motivational

structures. Whether it is right for me to steal a loaf of bread may depend, not just on whether I do or do not have a starving family to feed, but also on whether I accept the foundations of property norms, or relevant constraints on interacting with those who accept various property norms. My evaluative attitudes may play a crucial role in determining the rightness or wrongness of my action. If this can be so, then content universality can be separated from content specificity and the authority of obligation. Therefore, these kinds of relativisms do not undermine the objectivity of practical normativity and morality all together. The kind of objectivity I try to establish escapes the force of these kinds of diversity and relativism arguments.

Another kind of moral relativism argument which is meant to attack the objectivity of morality all together by first aiming at the content is J.L. Mackie's "argument from relativity" according to which, "the actual variations in the moral codes are more readily explained by the hypothesis that they reflect ways of life than by the hypothesis that they express perceptions, most of them seriously inadequate and badly distorted, of objective values" (Mackie 1977, 37). Mackie's version of the relativism argument from cultural diversity will not work if there are independent, well-supported arguments for a system of morality that would produce or allow for diversity. The strength on which the "better explanation" claim is rested derives from the portrayal of the alternative as misinterpreted perception. Without the misinterpretation clause, diversity does not provide the kind of evidence required by the argument. Thus, Mackie's argument really depends on a background claim that there are no strong arguments for the relevant kind of moral system. If we had a direct explanation of morality, especially one that allowed or even implied diversity of authoritative norms

among cultures and other groups, then this kind of relativism argument would carry no weight.

Whether or not I have presented the best versions of relativism arguments targeting content, there is good reason to think that these problems with the content of morality can be solved. Moral theories in general seek to provide answers as to the source of moral content and typically do so without appealing to much that is obviously culturally biased. Utilitarianism appeals to the universal desirability of happiness. Kant's ethics appeal to the coherence of one's principles with one's conception of oneself and one's conception of right action. Contractarians appeal to agreements we do or would make in order to serve our own purposes and live cooperatively. While these sources may only provide the beginning of moral content, they certainly go some distance in throwing doubt on the arguments which attempt to show the impossibility of specific or universal moral content.

In addition, there is little in the way of forceful argument that principled content is impossible because our beliefs and attitudes must arise in a particular cultural, sentimental or moral context. While diversity of moral code certainly exists, it is no proof that there cannot be any legitimate content to right or wrong actions. Of course, this is not to say that there are no reasons for being skeptical about moral content. There certainly are some reasons to be skeptical in this area. Rather, this is to say that the cause for concern is less philosophically demanding than it may initially seem, and less philosophically demanding than the problem of normativity.

Philosophers from David Hume to Charles Stevenson to Gilbert Harman have focused their concern on how there could be any real bindingness or normativity to our

decisions at all. The heaviest burden seems to fall on the issue of how our choices and actions could have the form of rightness or wrongness at all, regardless of how they should achieve any particular character of right or wrong in accordance with some rule. The problem is one of normative metaphysics and epistemology. It is a problem of how there could be the makings of right and wrong and, if there were, how we could know about them. It is not as much a problem of there being no contenders for universalistic normative principles. The problem of normative metaphysics and epistemology is where the analogy with the synthetic *a priori* applies. The problem with the synthetic *a priori* is how such knowledge could exist at all, rather than how we arrive at the particular knowledge we do. To solve the synthetic *a priori* problem, one might, for example, follow Kant in making room for a *form of intuition* (space and time) and *pure concepts of understanding* (the categories) which allow for synthetic *a priori* knowledge. The alternative of trying to solve the problem by showing, for example, how we know that three-dimensional space is the real space would be rather odd.

Of course, whether it is true or not that normativity is more in need of explanation than content, one cannot very well pretend to have achieved much in the way of explaining *morality* if one does not explain both normativity and, to some extent at least, the content – not to mention other important features of morality like social accountability. This follows from what has already been said relating normativity to specificity. Normativity requires specificity, because it is meaningless to say that I should without there being anything that I should do. Thus, to make headway in explaining morality, the reality of obligation and content specificity must be established. If this can be done without also establishing content universality, then the greater part of

the mystery will have been dealt with, and so the lack of established universality should not be considered a failure. I will deal at length with the idea of explaining morality without explaining content universality in Chapter 5. Much of the weight is taken off of the issue of universality by the stabilizing role of social accountability, which I introduce in Chapter 4 and elaborate on in Chapter 5, where this discussion is integrated with discussion on various threatening forms of subjectivism.

III. Constructive Explanation

So far, I have argued that skeptical attacks on objective morality show, not that there is no objective morality, but that objective morality needs an explanation. Further, I have argued that in order to treat morality with any kind of theoretical seriousness, we need to explain the normative aspect of morality – we need to explain how practical normativity fits into it. For now, this can be considered an explanation of moral obligation. To explain moral obligation and solve the mystery of how morality is possible, it is necessary to give a "constructive explanation."

> A *constructive explanation* explains a thing X by citing antecedent phenomena which non-causally produce X if juxtaposed in the appropriate way.

Intuitively, constructive explanation can be thought of as explanation by component parts. This definition must be further elucidated, of course. Phenomena are *antecedent* to X just in case they do not depend on X for their existence. One might think that this will not allow for the phenomena to be explanatory since they produce X and, thus, cannot exist without X also existing. If, for example, we were worried about how

there could be such thing as toast, because crunchy, starchy food seemed real but did not fit well with our preferred metaphysical-epistemological system, we might appeal to lightly charred bread as a possible explanation. The existence of lightly charred bread could perhaps be an explanation of how it is that there is such a thing as toast. However, one might well think that lightly charred bread cannot explain the existence of toast, since lightly charred bread cannot exist without toast also existing. The problem would be that 'lightly charred bread' is just a different way of denoting toast and so cannot be any kind of explanation of it.

However, phenomena that must coexist with X when properly conjoined need not coexist with X when these phenomena exist separately from each other. So, if we are worried about how to explain the existence of cars, we might appeal to wheels, axles, motors and steering devices. While wheels, axles, motors and steering devices can exist without an automobile coexisting with them (when they are not put together), when properly conjoined, there must also be an automobile. Thus, wheels, axles, motors and steering devices are antecedent to cars. One can cite the existence and conjunction of certain parts in order to explain the existence of a car and how it is that it exists and functions as a car. Similarly, charred food and bread can exist separately without there being any toast, but when properly conjoined, when the bread is lightly charred, there must be toast. Generally speaking, this may not be the most interesting form of explanation for the existence of a car or for the existence of toast, but it is a kind of explanation nonetheless. To take a different kind of example where the target is more questionable and in need of explanation, but such an explanation is probably unavailable or even impossible, we can consider what a constructive explanation of ghosts would be.

To explain ghosts, we would need to put together something that has crucial features of what ghosts are supposed to be. We would need to describe patterns, processes, etc., that exhibit "ghost-ness," that is, a mind, some empirical causal power (like the ability to make sounds, reflect a minimal amount of light, and perhaps cause larger objects to move), and penetrability. If a mind could be instantiated by patterns of activity that can freely move from one substance to the next, can create further motion in its surroundings, and so on, then explaining ghosts might be possible. If these features of ghosts are incompatible or physically unavailable, then ghosts cannot be constructively explained. All of these explanations appeal to the parts of the target object – constructive explanation is explanation by parts.

The set of phenomena Y *non-causally* or *constitutively produces* X if and only if the proper juxtaposition of the phenomena in Y constitutes the existence of X. While wheels, axles, motors and steering devices are antecedent to cars, we can also say that, roughly speaking, these things non-causally produce cars, since if they are put together in the right way, a car must coexist, they constitute the existence of a car. The proper juxtaposition of wheels, axles, motors and steering devices does not cause the existence of the car in any meaningful way. We do not want the phenomena to merely exhibit the less direct production relation of causation with the explanandum, because the point of a constructive explanation is to tell us what the explanandum is made of conceptually. A causal explanation would tell us about how things relate in the actual empirical world, rather than how they could be related at all.

The definition of 'constructive explanation' I am using requires that, in a sense, X is explained by appealing to the functioning parts of X. Antecedent, non-causally X-

producing phenomena are roughly equivalent to parts of X for the following reason: There must be multiple phenomena which jointly produce X and which could exist independently of each other. If this were not the case then they would all have to exist together and so they would have to coexist with X, since they jointly produce X, which would mean that they could not be antecedent to X. Multiple, independent phenomena which jointly must coexist with X (rather than causally lead to X) are, for all practical purposes, parts of X.

We do not want a causal explanation of practical normativity and morality, because that would not tell us how normativity works and how it is what it is. That is, a causal explanation, like a cultural-evolutionary explanation of moral practice does not tell us how morality works, unless it is presupposed that moral practice provides a constructive explanation of morality. In order to say just what morality is, how it fits into our world and how it has the qualities it is supposed to have, we need a constructive explanation.

In the case of explaining morality, a constructive explanation will require appeal to what is produced by how we think, choose and act. Appeal to these phenomena is required, because ultimately our thinking, choosing and acting embody our experience of morality, and our experience of practical normativity. To meaningfully explain a mysterious phenomenon, we must explain it in a way that makes use of how we experience it. Unless we do this, our explanation does not reconcile the problematic relation between our understanding of the world and our perceptions or intuitions about the phenomenon in question. Thus, for the mystery of morality to be solved and a constructive explanation of practical normativity and moral obligation to be given,

practical normativity must be explained by appeal to that by which we experience it and morality all together: our thinking, choosing and acting.

IV. Classical realism unsatisfactory

A. Classical Realism

One approach to accounting for the idea of moral and normative objectivity is encompassed by *realism*, according to which objective morality is built into the structure of the world. For classical realists such as Samuel Clarke, Richard Price, G.E. Moore and Peter Railton, rightness or obligation is a question of what actions *are* independently of how we think of them. As they see it, whether morality is real is the same as “whether *right* and *wrong* are real characters of *actions*, or only qualities of our *minds*; whether, in short, they denote what actions *are*, or only sensations derived from the particular frame and structure of our natures” (Price 1757, 1969, 657). About two hundred and fifty years later, Russ Shafer-Landau gives a slightly more refined, but very similar definition of realism. He puts it as the view, “that there are moral truths that obtain independently of any preferred perspective, in the sense that *the moral standards that fix the moral facts are not made true by virtue of their ratification from within any given actual or hypothetical perspective*” (Shafer-Landau 2003, p. 15, author's italics). According to this view, although the agent of the action must think about the action so as to deliberatively choose it, this is only required to make rightness applicable to the action – to make it an *action*. It is not required, in the sense of contributing to what rightness *is*.

I will not attempt to explore any of the nuances of the various varieties of moral realism. Perhaps most importantly, I want to point out that it is one of my basic starting

points of this project that normativity is to be accounted for and explained as something peculiar and mysterious. Because of this, I will not consider *ethical naturalism* – the view according to which "moral facts are a species of scientific facts, discoverable in all the ordinary ways, as motivating and as normative (or not) as ordinary facts" (Shafer-Landau 2003, p. 55). Ruling this out, this leaves moral facts and their normativity *sui generis* and, with regard to non-moral properties, simple. As such, I will restrict my use of the term 'classical realism' to the version of realism that has moral properties as *sui generis* and simple.

B. Classical realism cannot provide a constructive explanation of morality

The classical realist approach is metaphysical in that it increases the metaphysical complexity of our worldview by appealing to simple *sui generis* properties. This approach rules out the possibility of a constructive explanation of morality. To constructively explain morality one would have to provide phenomena, antecedent to moral obligation, which non-causally produce moral obligation when set together in some particular way. This, in a sense, would be to explain moral obligation by appeal to component parts. However, according to classical realism, moral obligation is *sui generis* and simple. This means that it has no parts. Moral obligation, on such a view, has no antecedent phenomena which non-trivially produce it. If it did, it would not be *sui generis* and simple.

Of course, it does not follow from the impossibility of a constructive explanation on the classical realist system that classical realism is false. The above argument is not a direct argument against classical realism. However, this does show that classical realism cannot help to solve the problem. It is asked, "How is morality possible?" to which

classical realism can give no answer and is even incompatible with the possibility of there being an answer in the sense of a constructive explanation.

While the classical realist may succeed in a sense in showing that there is no contradiction or absurdity in holding that rightness or fitness of actions is a simple, *sui generis* property, the account the realist gives is highly unsatisfying and not extremely helpful, since it is entirely incapable of producing the kind of understanding-increasing explanation of morality we seek. We are hoping to expand our knowledge and improve our understanding of the world, but when it comes to morality, classical realism cannot do this for us.

It seems that a helpful analogy can be drawn with a case from the natural sciences. Newtonian mechanics cannot be said to have ever really been absolutely and irretrievably contradicted by the empirical facts. However, it did reach a point where it became unable to produce new levels of understanding with respect to its new competitor, Einsteinian mechanics. In order to try to make progress, Newtonian mechanics had to be set aside in favor of Special and then General Relativity. Of course, the ability of General Relativity to succeed in explaining phenomena and develop our understanding can be taken as important confirmation for the theory, but the point remains that Newtonian mechanics ran into real trouble when it developed a kind of impotence rather than when it was found to result in contradiction or absurdity.

Similarly, classical realism should be set aside in favor of other approaches which present the possibility of expanding our understanding of moral phenomena. While classical realism may face no final contradiction or absurdity, it can never get around the impotence it leaves us to give constructive explanations and to develop our knowledge.

We want an integrated explanatory worldview. Classical realism puts limits on this. If we want to understand how morality is possible, and if morality must be understood in terms of how we experience it, then we must look elsewhere and leave classical realism behind.

The alternative to classical realism is to adopt and pursue a form of Kantian constructivism. The constructivist way is to try to explain moral obligation, and with it morality, by appealing to what we create by our thinking, choosing and acting. These features of our experience involve both motivation and normativity, and they embody our experience of morality altogether. For the purpose of explaining morality, these phenomena can serve as the components of a constructive explanation. Thinking about actions, choosing actions and acting are phenomena antecedent to moral obligation in the appropriate sense. If it can be shown that they produce moral obligation, when more detail is provided and they are appealed to in a more restricted and refined way, then this will be the basis of a constructive explanation of morality. This is the Kantian constructivist approach to the foundation of moral obligation.

V. Obligation by agreement unsatisfactory

One common strategy for trying to explain morality is claiming that it arises from some explicit, implicit or hypothetical contract or agreement. On such a view, morality can be seen as a way of solving the problems of conflicting interests in a world where almost everyone does or would want to cooperate with each other, under the right circumstances. This approach is taken by Hobbes and Rousseau, and more recently by Rawls, Gauthier and Scanlon. The theories of these philosophers differ greatly and the

way that they understand the nature of the prior, continuing or hypothetical agreement is different in each case, but they all share in the idea that morality, something like morality, or a replacement for morality arises from some form of agreement.

If we take these theories as attempts to explain morality, then we can ask, on the one hand, whether they can explain features of morality like accountability and content, and, on the other hand, whether they are capable of success at accounting for practical normativity and obligation considering the tools they make use of. Assuming the legitimacy of some kind of obligation or normativity arising from, or acting with the given type of agreement the theory appeals to, these kinds of theories seem to have the potential for providing specific content, perhaps universal content, and maybe social accountability. The works by the above mentioned authors and others have gone a long way in showing the fruitfulness of the contract-based approach in generating moral content and accountability by what people do or would (under some specified set of important circumstances) agree to. I take it that the bindingness of the right kind of agreements would have the potential to substantiate the rightness of many of the actions typically taken to be morally obligatory in everyday life. Various forms of agreement can explain how it is that we have the cooperative obligations we think we do, assuming the normative authority of such forms of agreement. Social agreement can also explain why it is that we are accountable to one another for compliance with the norms of social living that we take to be authoritative, on the assumption that there is some appropriate authority to the social agreement itself. This, in fact, is the route I take in filling in moral content and social accountability. By explaining a general scheme of normativity, and then explaining how this scheme of normativity supports the authority of the right kind of

social agreement, I can explain moral content that promotes cooperative living and social accountability as an integral part of our obligations. In Part III, I elaborate on a role for contractualism within my theory of moral obligation. However, at present, the capacity for such theories to explain moral *content* on the basis of some form of agreement is not the point I want to press concerning their overall potential to explain morality, and so I will not discuss it any further. What I currently want to examine is whether agreement theories can explain the normativity of obligations without any assumption of prior normative authority.

The main problem with taking such theories as explanations of morality all together, as I see it, is that agreement itself does not explain the reality of obligation, or the normative aspect of obligation, whether the agreement be actual or hypothetical. What obligates me to do what I agree to? It certainly does not follow logically from the fact that I agreed to do something, or that I would have agreed to do something that I am obligated to do it, and that the agreement is authoritative for me. There seem to be three possible ways of filling in an account of obligation in a contract-based theory of morality, each of which fails to explain the normative aspect of obligation on its own.

The first possibility is that the theory relies on an implicit classical realism. It might be claimed that it is just a brute, inexplicable fact that it is right to do what one has agreed to do, or would agree to do under the right circumstances. This is one way of reading a line of argument in Hobbes (Hobbes 1651, 1968). One of Hobbes's arguments is that we agree to live by the rules of society and what we agree to we are bound by, because going against what I have agreed to is like contradicting myself: it is an absurdity. This can be interpreted in a classical realist way whereby Hobbes takes the

obligation of agreements to be brute and inexplicable (Taylor 1938, 406-424;Warrender 1962, 434-449;Rhodes 1990;Harvey 1998), although he cannot say of course that it actually does follow logically. As we have already seen, this does not suffice to explain obligation, because it does not explain the normativity. This does not solve the problem of how morality is possible in the form in which it is here being considered.

The second way the contract-theorist can fill in the explanation is by appealing to the instrumental power of the alleged contract. This is the approach of David Gauthier (Gauthier 1986) and is a standard way of reading Hobbes's main line of argument, and involves claiming that adherence to the contract is the best or only way to achieve goals we happen to have anyway. Abiding by the actual or hypothetical agreement is a means to our end of a more comfortable or secure life. This approach does not explain the reality of obligation, but rather does away with it and substitutes the motivational force of explicit ends. As a general theory, there are many problems with this view, some of which spring from the shedding of obligation, such as the free rider problem and its inability to provide a way of ordering one's own ends in terms of priority. Rather than try to show that such a theory cannot work, I will simply point out that this cannot be an explanation of practical normativity or of moral obligation, and so it cannot solve the problem of how morality is possible. Whether this type of theory can be taken as evidence countering the appearance that morality is possible is not an issue I will pursue.

The third, and perhaps most important way to understand an explanation of obligation in terms of agreement is with the notion of "practical contradiction." For my purposes, *practical contradiction* is violation of one's own adopted rule. If I adopt the rule for myself that I do not cross the street when I have a red light, then I practically

contradict myself if I do cross on red. This is another way to interpret the argument from Hobbes mentioned above to the effect that going against what I agreed to is a kind of absurdity. This differs from the classical realist version of obligation by agreement in that agreements, under the right conditions, are interpreted as adoption of self-governing rules, and this is held to be binding. The realist version holds that agreements are just agreements or perhaps even promises, and that these are fit or right to be followed through as a brute fact about agreements or promises. Practical contradiction is a step in the right direction. It brings us closer to explaining the force of the obligation stemming from agreements.

However, practical contradiction also fails to explain the bindingness we are looking for. We still need to know what it is that makes violating one's own rules wrong and following them right. We might assume that the particular contract theory will explain what it is about the particular rules we adopt by way of the agreement that makes it impossible, in some sense, for us to just change our rules in order to get out of their bindingness. If there is no such restriction, then just as Hobbes claims is the case for the sovereign, the rules are not meaningfully binding because we can always just drop them whenever we want. For this account to work, there needs to be something that keeps us from changing our rules arbitrarily, but there also must be something that makes violation of one's own rule wrong. Without anything showing otherwise, there seems to be nothing fundamentally different about violating a rule that I made as opposed to a rule someone else made.[1]

---

[1] I am not claiming here that one's own rules are not practically binding when one does not change them. I will later argue that one's commitments, until they are changed and reorganized, *are* normative for every rational agent. However, my point here is that this needs an explanation, and cannot be considered an explanation of practical normativity. In Part II, I will argue that it is the structure of rational agency which

So, agreement theories, by themselves, cannot explain moral obligation. Although they can explain the content of obligations and social accountability under the assumption that the appropriate agreements are in some way binding and authoritative, they cannot explain that very bindingness and authority of the agreements. Later, I employ a version of agreement theory to do the work of explaining social accountability and the content of moral obligations. But, another kind of theory is needed to explain a general scheme of practical normativity that can give authority to the content and accountability relations that make up our moral obligations. This combination may be very difficult, however, because we do not know in advance that the two theories – the theory of normativity and the agreement theory – will fit together cohesively. It may be that explaining a general scheme of practical normativity creates a problem for adopting a socially-oriented theory of moral content and accountability. I discuss this issue at great length in Chapter 4.

VI. Constitutive internalism

So far, I have argued that we need an explanation of the force of normativity and how that force can be fixed to content at all. This stands opposed to an explanation of where the particular, actual content of what is normative or obligatory for an agent comes from. I now want to discuss further constraints on an explanation of practical normativity, which would answer our question of how morality is possible. I want to

makes commitments binding on an agent. This requires a complex explanation and is a far cry from simply claiming that one is bound by one's own rules. The argument draws from the nature of agency, its conceptual role in action and deliberation, and the constitutivist strategy for explaining practical normativity, all of which I develop in Chapter 2.

look at the role that motivation and the practical structure of the agent must play in such an explanation.

A lot of fuss has been made about “internalism” and motivation in the literature and how that might figure into moral obligations, practical normativity, and the coherence of classical realism (Darwall 1992, no pages; Darwall 1983; Audi 1997; Shafer-Landau 2003; Prichard 1928, 2002; Falk 1948, 111-138; Nagel 1970; Williams 1981, 101-113; Korsgaard 1986, 5-25).

In this section, I want to look at some different ways in which motivation might be connected with normativity and obligation. I do not discuss any of the standard arguments for why internalism must be true or for why it must be false. Instead, I want to connect it with the issue of giving an explanation of normativity, and how this relates to the kind of moral theory we look for. I argue that to explain normativity and moral obligation in a substantial and interesting way, we need to do so through a version of the view that motivation is a constituent part of what it is to have an obligation.

Further, I relate this to classical realism by looking at the relations between realism and relevant kinds of “externalism,” the denial of the claim for a necessary role of motivation in obligation, and what motivations there are to accept these views. I set up as a parameter for an explanation of moral obligation accepting constitutive internalism and appealing to the practical-motivational structure of the agent, but it can also be seen as a way to thwart the background motivation for classical realism. Thus, there is a kind of harmony between the Kantian line of explaining morality through the structure of deliberation, agency and action, and letting go of the independence of moral properties from what arises from our motivation.

A. Kant and Falk

In an important paper, "Ought and Motivation" (Falk 1948, 111-138), W.D. Falk made a distinction in ethical theory which has many consequences for the discussion of the existence and nature of moral obligation. In response to H.A. Prichard in "Duty and Interest" (Prichard 1928, 2002), Falk described and argued for a position he called "internalism" maintaining a necessary connection between having an obligation to X and having a motive to X. In this section, I describe the various facets of this and related positions and discuss the relevance of internalism in general, and Kant's form of internalism in particular for the overall project. The way that Kant handled the relation between obligation and motivation is of vital importance for my project, as it is the view I adopt. On this view, having a motive to X is part of what it is to be obligated to X. Motivation is a constitutive component of obligation.

As an apparent solution to a practical problem in which many people seem to find themselves, Prichard presented an analysis of the relation between duty and motivation. He maintained that an individual who recognizes an obligation but still asks for a motive to perform it is overlooking the fact that an individual is always motivated to perform duties when he or she recognizes them, because it is a fact of our moral psychology that we have motives to do what we take to be our duty.

Falk denied the acceptability of this analysis on the grounds that it mishandles an apparent fact about the connection between 'ought' and motivation. According to Falk, when I accept that I ought to X, I am at the same time accepting that I have a motive to X, which entails that I could X and would X if nothing got in the way. I must have a disposition to X in order to have an obligation to X, and not merely in order to *recognize*

that I have an obligation to X, because obligations come from our reasons and reasons are motivational. According to Prichard the connection between obligation and motivation is psychological, while according to Falk the connection is logical.

It is a crucial aspect of Falk's position that having a motive to X is not equivalent to having a desire to X. Motives, according to Falk, are thoughts which impel a person toward action when attended to. A person can have motives "in either an *occurrent* or *dispositional* sense; for 'he is being impelled (or caused) by the thought of some act to do it', or 'he would, if he dwelt on it, be impelled (or caused) by the thought of some act to do it'" (Falk 1948, 25). In neither of these senses is having a motive equivalent to having a desire. Having an occurrent motive means *being* impelled to do some act, whereas desiring means *feeling* impelled to do some act. Thus, having a desire means "we are impelled and that we have perceptual evidence of our being so" (Falk 1948, ibid.). From this it follows that "To say that someone desires to do some act would therefore go beyond saying that he had a motive or was impelled, even if, whenever anyone were impelled to action, he also felt desire; but in fact people are often impelled to act under the influence of a motivating thought without noticing any desire at all" (Falk 1948, ibid.). From this we can extrapolate that Falk's internalism does not imply that to have an obligation to X one must desire to X.

In the case of Kant, it is clear that he was an internalist in Falk's sense. In the *Groundwork*, Kant claims, "morality is a law for us only as rational beings" (Kant 1785,1911, 447),[2] which means that having an obligation implies being rational. In the *Critique of Practical Reason*, Kant writes, "Pure reason is practical by itself alone, and

[2] All references to Kant's works are from the Prussian Academy editions. Many translations have the page numbers from the Prussian Academy edition in the margins.

gives (to man) a universal law which we call the moral law" (Kant 1788,1913, 31, orig. pg. 56). Thus, the moral law comes from reason and reason has the power to determine actions. From this it follows that being rational implies capability of moral action, which implies the existence of a motive to perform moral action. Therefore, for Kant, having an obligation entails having at least a dispositional motive to perform the obligation.

Having established that Kant was an internalist in Falk's sense, which is important for the discussion in what follows, let us now consider what obligation is according to Kant. For Kant, morality is what I do insofar as I am rational. Morality is just accordance with the categorical imperative, "Act only on that maxim which you can at the same time will as a universal law" (Kant 1785,1911, 421), which Kant claims is exactly what beings do to the extent that they are rational (Kant 1785,1911, 440). That I *ought* to do something, or that I am *obligated* is restriction to an action or *necessitation* in the face of what reason would have me do (Kant 1785,1911, 413)(Kant 1788,1913,32,orig. pag. 57). As a being whose rational will does not always determine action in accordance with reason, or *autonomy*, "the property of the will, through which the will is a law for itself" (Kant 1785,1911, 440), I do not always do what reason directs. As far as reason or autonomy directs my actions and has content as to what would be done in accordance with it, I ought to do that and I am obligated to do it. I am obligated to do X just in case I would do X if my actions were determined by reason.

B. Distinctions in internalism

Following Stephen Darwall (Darwall 1992), we can distinguish between some different forms of internalism. While Falk only set apart the view that having a motive is entailed by having an obligation, there are a few different, but related views about a necessary connection between obligation and motivation which can be distinguished.

Perhaps the most commonly discussed form of internalism is *Judgment internalism*, "the view that sincere *assent* to [an ethical] judgment concerning what one should do is necessarily connected to motivation" (Darwall 1992, 155). This is the view that Hume espouses when he says "morals . . . have an influence on the actions and affections," and, "Morals excite passions, and produce or prevent actions," where by "morals" Hume means "the opinion of . . . obligation" (Hume 1739,1978, 457). This is also the view often associated with expressivism and related theories of metaethics, which assert something like the claim that sincere assent to an ethical judgment just is a kind of motivational state.

While judgment internalism may often be used as a way of getting rid of what one *in fact* ought to do, other forms of internalism move away from this tendency. *Existence internalism* is the view that "someone [ethically] ought to do something only if, necessarily, she (the agent) has . . . motives to do so" (Darwall 1992, 155). Darwall considers this a "metaphysical" view in that it makes a claim about what must hold in order for there to *be* an obligation, rather than what must hold in order for someone to *believe* there is an obligation or have the *attitude* of present obligation. Falk certainly asserted this, although he also thought it could be argued that more follows from it, by claiming that a merely psychological connection could not support a necessary connection.

Existence internalism, which allows for moral facts, may be further divided into constitutive and non-constitutive internalism. *Constitutive internalism* is the view that "motivation is a constituent of ethical facts themselves" (Darwall 1992, 157) and not just an effect of the recognition of ethical facts. This was Kant's view and it was also Falk's

view. According to them, motivation in some form is part of what it is to have obligations. Insofar as one has obligations, they must be based on reasons which must spring from our motivations. This is the view I endorse. My explanation of practical normativity and moral obligation comes as a form of constitutive internalism.

Non-constitutive internalism is just the complement of constitutive internalism. This view was held, for example, by Richard Price (Price 1757, 1969, 655-762) and H.A. Prichard (Prichard 1928, 2002). They claimed that what duties are has nothing to do with the agent's motivational states, but that, for some reason beyond what duties or obligations are, agents are in fact necessarily motivated to perform their duties when they recognize that they have them. There is some part of our moral psychology that produces motivation to perform our obligations whenever we realize that we have them. Falk did not acknowledge that this was a form of internalism (Falk 1948, 111-138), because he argued directly against Prichard that there could not be such a necessary connection. Any relation between motivation and obligation would have to be contingent, he thought, if having a motivation has nothing to do with what it is to have an obligation. If the two are separable, then they are not connected necessarily.

Existence internalism is often paired with the view that there are things people ethically ought to do and that in all cases those people are necessarily motivated to do the things in question. However, one might accept existence internalism and take it to show that there cannot be any universal moral imperatives, since, one could argue, there exist no actions for which all agents have motives. This can be countered by the contention that duties do not need to correspond with some independent motives, but that duties *produce* motives by the universal connection between the thought of duty and motivation;

or by claiming that the very facts which provide for us having duties are those which provide for us having motives.

C. Internalism and externalism

To continue in the Kantian vein, I propose a solution to the problem of how morality is possible with a version of constitutive internalism. To help motivate this move, and to show how constitutive internalism is a natural way to pursue a constructive explanation of morality, I discuss the relation between classical realism and externalism, and the motivation for realism and externalism. In order to do that, I first introduce a distinction between two views which will serve as intermediaries between classical realism and the different externalisms and internalisms defined above.

Explanatory internalism is the view that normativity or obligation requires a contribution *from* the functioning of the deliberative system of the agent *to* a constructive explanation of obligation. *Explanatory externalism* is the negation of explanatory internalism: that normativity and obligation do *not* require a contribution from the functioning of the deliberative system of the agent to a constructive explanation of obligation. These definitions are clearly similar to those given by Darwall of existence and constitutive internalism, but they are conceptually distinct. These notions of internalism and externalism will be helpful in mapping out the relation between classical realism and the negation of constitutive internalism.

Although I have defined explanatory internalism differently than constitutive internalism, these views turn out to be effectively equivalent. Constitutive and explanatory internalism are equivalent, because, first, normativity is a constituent of ethical facts, and explanatory internalism holds that the functioning of the deliberative system, which is motivational, must contribute to normativity, that is, be a constituent.

So, if the functioning of the deliberative system must contribute to a constructive explanation of normativity, then motivation is a constituent part of obligation. So, explanatory internalism implies constitutive internalism. Second, there is no other way for motivation to be a constituent of ethical facts than for it to contribute to the normative constituent. And, there is no other way for motivation to contribute to normativity other than via the functioning of the deliberative system. So, if motivation must be a constituent part of obligation (or ethical facts), then the functioning of the deliberative system must contribute to a constructive explanation of practical normativity. So, constitutive internalism implies explanatory internalism. It follows from this that the negations of these views are equivalent: explanatory externalism is true if and only if either existence externalism or non-constitutive internalism is true (constitutive internalism is false).

We can also see that classical realism implies explanatory externalism. According to classical realism, normativity is simple and *sui generis*. Thus, there is no room for a contribution from anything including the functioning of the deliberative system of the agent. In a certain restricted sense the converse is also true. If explanatory externalism is coupled with claims to the effect that there are objective moral properties of the right kind, this still does not imply classical realism, but it does seem to leave it as the only interesting or tempting option. If rightness is not simple and *sui generis,* and so there are constituents of normativity, and none of them derive from the functioning of the deliberative system of the agent, such phenomena do not lend themselves to description. As far as we would be concerned, they might as well not exist and rightness might as well be simple.

What is particularly interesting about this connection between classical realism and externalism is that classical realism gains much of its motivation from externalism. Realism is motivated by externalism in that if one assumes externalism, one may easily think that since obligations must arise independently of any internal state in the agent, they must be explanatorily basic with respect to the functioning of the agent's deliberative system. As argued above, in a restricted sense, this implies (or leads to via plausible reasoning) classical realism. One is led to classical realism by the thought that obligations must be independent of motivational states (externalism), and thus obligations must be simple and *sui generis* (realism), since there is no reasonable alternative for how obligations could be complex, but not built out of motivational states in the agent. This reasoning is unaffected by consideration of the claim that, although it has nothing to do with what obligation is, one is necessarily always motivated when one recognizes a duty (non-constitutive internalism).

If externalism is the primary motivation for classical realism, then we are led to wonder what the motivation for externalism is (the negation of constitutive or explanatory internalism). Explanatory externalism seems to be motivated by the thought that duties must exist independently of the agent's motivation, otherwise there is no force to the obligation. In other words, the bindingness of a duty seems to be illusory if it can only exist where one already has motivation to perform the act. Thus, the reality of obligation altogether seems to require explanatory externalism.

Obligation carries with it a kind of force, authority or necessity which gives it its unique normative character. In order for obligation to have force on, or authority over the agent's will, it must exist apart from the agent's will. That is, it might seem

implausible (or impossible) that such force, authority or necessity should depend on what is inside the agent. All that is required, it might seem, is that the individual (the agent) can grasp practical force, authority or necessity in general. Whether it applies in a particular case should not depend on the particular state of the individual in that case. If it did depend on the state of the individual in this way, then there could be no force or authority, since it would seem that such a phenomenon must compel the agent with indifference in order to exist. The failure of obligation to track the agent independently of the agent's motives is for the obligation to ride on the agent's mechanistic psychology, which is no obligation at all.

This last step in the argument is the crux of the matter. The motivation for externalism boils down to the claim that there is an exhaustive dichotomy between externalism and subjectivism or the unreality of obligation. If I can show this to be a false dichotomy by providing an account of obligation, wherein the force and authority does depend on the individual's particular motivational state in the particular case, *and* the force really does exist, then this motivation for externalism will be defeated. This means that the motivation for externalism and classical realism, if I am accurately representing them, is really just skepticism about the possibility of deriving obligation from an internalist framework.

Therefore, although my primary goal in this essay is to give a constructive explanation of moral obligation, if I can achieve it, I will at the same time be achieving a secondary goal of showing the falsity of the externalist-subjectivist dichotomy, and thereby undermine a great part of the externalist, and therewith classical realist motivation. In order to do this, I give a constructive explanation of moral obligation by

analysis of what it is and what it entails to be an agent, and taking various internal motivation-laden states which are antecedent to moral obligation, and showing that they produce moral obligation.

Chapter 2
The Constitutivist Approach to Practical Normativity

I. Introduction to Part II

In the last chapter I argued that morality needs an explanation. I gave the further analysis that the most pressing need was for an explanation of practical normativity. In this and the next chapter, which make up Part II, I attempt to provide the desired explanation of practical normativity. In Chapter 1 I also argued that to truly explain practical normativity one has to do so with an internalist theory, that is, the explanation would have to make the agent's motivation a component part of practical normativity for that agent. As such, my proffered explanation makes the agent's own motivation the very foundation for practical normativity. The theory I offer cannot be faulted for failing to meet any internalist requirements. This is a large part of why the theory I offer is so thoroughly explanatory.

A rough sketch of the explanation I give can begin with a general take on evaluative attitudes. Like Sharon Street argues, we can account for one's actually having a reason to φ in terms of what one takes oneself to have a reason to do – in terms of one's evaluative attitudes (Street 2008). For example, if I take myself to have a reason to survive, then I have a reason relative to this attitude not to drink poison. Further, if my complete set of evaluative attitudes directs me (by some measure) not to drink poison,

then I should not drink poison. We do in fact take ourselves to have reasons for things, or in other words, we do in fact have evaluative attitudes. So, it seems that there is at least this source of reasons that we actually have.

There are important constraints or parameters for how this works. First of all, one cannot really be operating on reasons if one is not trying to make one's evaluative attitudes jointly coherent and consistent. This is a common requirement and many philosophers seem to think it is obvious that coherence forms some sort of constraint on one's evaluative attitudes even if they cannot say exactly why (Street 2008;Gewirth 1978;Gibbard 1990;Smith 1994), although I not only claim this but give an explanation of how and why coherence forms a constraint on one's evaluative attitudes.

Another important constraint, however, is systematic unity. One cannot really be operating on reasons unless one is trying to make one's evaluative attitudes systematically unified. So, if one has the attitude that being good to one's pets is valuable, then one must be trying to incorporate that attitude into a network of evaluative attitudes that have some joining principles. For example, one might take it that all animals are worthy of concern and that taking ownership of an animal imparts a relationship of responsibility and care on the owner. These attitudes might tie together various other attitudes that the pet-lover held and could be further unified by more abstract attitudes. Trying to connect one's evaluative attitudes through systematic unity is part of what it is to be a rational agent and operate on reasons. This is similar to Michael Smith's view (Smith 1994;Smith 1997, 84-119), but I argue that it is systematizing our commitments (in a technical sense) that practical rationality requires, not our desires. Most important, however, is that I give a thorough-going explanation for

why systematization is a rational requirement, instead of simply appealing to the intuitive plausibility that it is a rational requirement, as Smith does.

The structure of agency and the systematization requirement also give us a standard of authority between one's various evaluative attitudes, which must surely come into conflict with each other on a regular basis. No one has a perfectly conflict-free set of evaluative attitudes. To use a common metaphor, those attitudes that are closer to the center of one's interwoven web of evaluative attitudes have more authority when conflict arises. To put it another way, those attitudes that bear more relations of systematic unity to the whole scheme of attitudes have more authority than those that bear fewer. My attitude that puts great value on protecting my family has many systematic relations and has much greater authority than my valuing of almonds. So, if I am in a situation where I must choose to uphold one or the other attitude, then I have more reason to protect my family than to eat or obtain almonds. If my wife needs my help to save her well-being, but it will require me to take a drug that will make me allergic to almonds, then I should go ahead and take the drug, despite the damage this will do to my almond-valuing.

The reason I end up with this standard of authority based on systematic centrality, is that it comes straight from what it is to act as an agent. To make a decision and perform an action as an agent is to draw from one's entire motivational field as a unity and make a single, encompassing decision, and then to act on that decision. Since this requires aiming at coherence and systematic unity among one's motives and attitudes, and the greatest coherence and unity is achieved by affirming those attitudes that are most central within the systematic scheme, affirming the central-most attitudes does the most to meet the standards internal to the practice of performing an action as an agent. This

means that affirming the central-most attitudes is a standard of authority for the deliberation of an agent, which is what is at issue.

Although this is my view in relatively intuitive terms with reasoning that, to some, may seem compelling enough, it is not an explanation of the kind we are after, because it leaves too much detail out and leaves too much to intuition. To really give an explanation, all of the subtle nuances must be brought out and elaborated. In particular, the following questions must still be answered: How can one account for the authority of reasons in terms of what one must do in order to operate on reasons or to perform an action as an agent? Why does agency require an attempt to make one's evaluative attitudes coherent and systematically unified? How and why does this supposed standard of authority among evaluative attitudes constitute a general scheme of practical normativity? What does it take for an agent to perform an action? What exactly can be considered attributable to an agent, and is there more than one sense in which an act can be attributed to an agent? Is there a particular form that the decision of an agent must take? Assuming the standard of authority and reason-giving based on systematicity is correct, how does rational deliberation work when done correctly? These and other questions must be answered to fill out my explanation of practical normativity.

The strategy I use to explain practical normativity is to argue that a certain aim is constitutive of agency and rational deliberation, and from it to argue that this provides a system of practical normativity because it is an aim,[3] and the aim has a constitutive role in agency and deliberation which determines what successful agency or deliberation is.

[3] Although Connie Rosati suggests, without committing herself, that a capacity might be part of a constitutivist theory, rather than an aim, I do not see how this would work. "Agency and the Open Qestion Argument," *Ethics* 113 (April 2003): 490-527. In what follows, I argue on behalf of the constitutivist strategy assuming that an aim of agency is what the analysis in question produces. A motive or inclination might be substituted, but this does not seem to add anything to the analysis.

This strategy, *constitutivism*, has been championed by J. David Velleman (Velleman 1989; Velleman 2000b; Velleman 2004, 225-238; Velleman 2004, 277-298) and Christine Korsgaard (Korsgaard and others 1996; Korsgaard 2002; Korsgaard 1999, 1-29). In their work they have each, separately argued for a constitutive aim of the core elements of practical nature, action and agency, and used these arguments to try to explain practical normativity. They have done this in distinct ways: Korsgaard has argued that the constitutive aim of action or agency is the unity of the agent. Velleman has argued that the constitutive aim of action is self-knowledge and conscious control.

I make my own constitutivist arguments which differ from both Velleman's and Korsgaard's. In the next chapter, I argue that systematization of one's commitments is the constitutive aim of rational deliberation. This is a concept I have to develop over the course of this chapter and the next in order to make clear what I mean by this. In this chapter, I focus mostly on setting up the constitutivist framework. First, in the subsequent section, I argue that constitutivism in general has the potential to succeed. I argue that the general formula of constitutivism has what it takes to explain practical normativity.[4]

In the rest of this chapter, I argue for the supporting premises of my particular constitutivist theory. I argue, for example, that rational deliberation is the appropriate point of focus for an analysis of right action, and that an agent causing his or her own behavior is the foundational concept of all rational action. I present an account of what has to happen for rational action to occur and what relationship that has to other forms of behavior we would consider intentional. Additionally, in order to present a preliminary

[4] While Velleman, in particular, also argues that constitutivism can explain theoretical normativity J. David Velleman, "On the Aim of Belief" In *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000)., I will not attempt any analysis of this and will restrict all of my attention to practical rationality.

version of my account of practical normativity, I argue and use a more general statement of my core constitutive claim concerning systematization of commitments. This is Korsgaard's claim that the constitutive aim of agency is unity. While her claim is different from what I espouse in Chapter 3, it can be considered an approximation of what I argue for. Bringing this in allows me to present a complete argument for a constitutive account with the framework that I set up in this chapter. This, in turn, allows for a preliminary look at the argument in a somewhat simplified version, which helps show how the argument works and what is at stake in Chapter 3.

So, in this chapter, I begin with a look at the argument strategy I am using – constitutivism. I then carry out this strategy by setting up a formal argument that would explain practical normativity. The argument is presented in full in Section III, in which I also discuss the meaning and significance of the premises. Thereafter, I defend each of the premises in turn, over the course of the rest of the chapter. In Chapter 3, I present the argument again, but instead of the unity interpretation of the constitutive aim, I present and defend my own analysis of the constitutive aim of agency and rational deliberation – systematization. It is the task of Chapter 3 to make and legitimate this change and to present my full and official explanation of practical normativity.

II. Constitutivism

A. The basic idea – normativity springs from constitutive aim

In this section, I have two goals, one general and one specific, where the specific goal helps meet the general one. The general goal is to present the constitutivist strategy and explain what it is aimed at establishing and how it works. The specific goal is to

respond to some very important objections to constitutivism from David Enoch (Enoch 2006, 169-198) and show them to be in error. Because these objections get right at the foundational points of constitutivism, discussion of them greatly helps in elucidating the constitutivist strategy and establishing its legitimacy. To serve my general goal, I lay out the skeleton of the constitutivist strategy and argue for its basic tenets and implications. I outline the constitutivist strategy and explain how a constitutive aim can govern by determining what counts as eligible for the application of the standard of right action or successful agency. Exactly where this governance over the rightness of action gets its foothold is argued for, and I analyze what follows from the core constitutivist ideas.

Thereafter, I examine the objections that Enoch makes and respond to them by pointing out how they are based on a misconception of constitutivism. Enoch has sought to show that normativity cannot be explained by what is constitutive of action, agency or rational deliberation. While I argue that his objections are in error, he does bring to light the crucial steps in the strategy, and, so, discussion of his criticisms is of great assistance in expounding constitutivism. Enoch's objections amount to the claims that constitutivism cannot explain a relevant sense in which an agent has a reason to pursue the constitutive aim itself, and that having such a reason is necessary for the aim to supply derivative normative reasons. I agree that constitutivism cannot explain a relevant sense in which an agent has a reason to pursue the constitutive aim itself, but I argue that constitutivism can explain reasons that derive from a constitutive aim without relying on any sense in which one has a practical reason to pursue the aim.

The idea that there is a constitutive aim of agency and that such an aim can be the ground of practical normativity all together is a nuanced and perhaps radical one.

Thematically, this strategy is Kantian in the sense that Kant thought normativity springs directly from the nature of rational agency. Kant could be read as a constitutivist, perhaps, but Kant never offered an explanation at the same level of detail in how the strategy is supposed to work which is now being brought forth by constitutivists.

The constitutivist line is that there are certain aims, which are essential to agency, and these aims provide reasons relative to the aims in just the mundane sense it is normally accepted that one can have reasons relative to an aim. For example, if I aim to cook new and interesting dishes, then I have reasons, relative to that aim, to cook things I have not cooked before and that, by some standard, appear interesting. If I aim to improve my physical fitness and health, then I have reasons, relative to this aim, to do that which is conducive to improving fitness and health. Relative to this aim, I have reasons to get more exercise and to eat better. I would have these reasons, if I had these aims, although the reasons would only be as stable and compelling as the aims that they are relative to.

For these reasons to cook things I haven't cooked before, to eat better and get more exercise to really be normative for me in the absolute sense we are looking for, there must be something more which backs up the normativity of following my aims to cook new and interesting dishes and to improve my fitness and health. These aims I have might be rather whimsical. I might have no reason to aim at getting healthier. Perhaps I am already healthy enough, or maybe health doesn't matter that much. In order for me to really have a non-relative normative reason to eat better and get more exercise, which stems from my aim to be healthier, my aim to be healthier must be grounded normatively somehow, or somehow the relativity of the reasons must be discharged.

In the case of the constitutive aim of agency, the thinking continues, one has reasons relative to an aim which all agents must have, which are, hence, reasons for all agents. In a certain sense, this already takes away the relativity of the reasons that stem from the aim of agency. This is an argument that Velleman uses to establish the authority of the constitutive aim (Velleman 1996, 694-726). Velleman makes his point in terms of internalism and externalism, that is, whether and in what sense there is a necessary connection between having a reason to φ and having a motive to φ. Internalists maintain that one cannot have a reason to φ without having a motive to φ, while externalists deny this.[5] As Velleman argues, reasons that come from an aim of agency come from actual motivations of the agent, and so a constitutivist account of practical reasons is a fully internalist account. But, constitutivism can also be said to give a kind of externalist account of reasons in that reasons that come from an aim of agency hold for all agents, and so are independent of the particular motivations of particular agents, because all agents must have the aim. Hence, reasons that come from the aim of agency are not relative to the particular aims of particular agents.

This way of discharging the relativity of the reasons may seem unsatisfactory, however, since it might seem to simply show that reasons following from the aim of agency are equally relative for all agents. But there is more to constitutivist theories than just the fact that reasons stemming from the aim of agency are equally relative or non-relative for all agents. By using an analysis of what agency is we can reveal what the standard for successful agency is, which in turn will be a standard of practical normativity. Since the aim is, according to a particular constitutivist theory, the aim that

---

[5] See Chapter 1, Section VI.

one has insofar as one is an agent, that is, it is the aim one has *as* an agent, it sets the standard for what successful or unsuccessful agency would be. Constitutivism uses the Kantian idea that we can and must explain practical normativity by starting with an analysis of the first-person, deliberative standpoint and working from there. If the first-person, deliberative standpoint can be found to have a particular aim, an aim that is constitutive of it, then we can see what constitutes success from this standpoint. And, what constitutes success from the deliberative standpoint is what is practically rational or what is practically normative.

If the appropriate arguments for the conclusion that there is some specific constitutive aim of agency are sound, they show why there are normative standards for choice and action, what they are (at least in the abstract), and when they apply. There are normative standards for choice and action, according to the constitutivist view, because there is an aim that determines what deliberation or agency is and various directives follow from the pursuit of that aim. The fact that the aim is constitutive of agency is what takes the relativity out of the reasons that derive from the aim. In general, aims give rise to reasons relative to them. The constitutive aim gives rise to reasons relative to it, but these must be reasons in an absolute sense, since they come from what is constitutive of being subject to reasons all together.

There are some interesting consequences of this view, which are very important later on when I address Enoch's objections to constitutivism. One corollary of a particular constitutivist theory is a specification of the limits of first-person, deliberative practical normativity. In taking up the deliberative standpoint, I am confronted with the need for reasons. I ask, "Why should I exercise?" or "Why shouldn't I eat the whole

pie?" This calls for first-person, deliberative normativity. When I ask these questions, I am looking for practical reasons that direct me one way or another on these issues. If there really is a constitutive aim of agency, then I must be aiming at that certain something whenever I try to answer these kinds of practical questions as an agent. But, if I do not have this special aim, then, whether or not I ask the question, "Should I eat the whole pie?" there is no real answer. There is no real answer, because I am not, in such a case, an agent. This means that I am not capable of having practical reasons. There might still be a sense in which you could sensibly say that, even though I am not an agent, I should not eat the whole pie because it is bad for my health or I will feel bad for it later and there are substantive standards that make pursuing my interests in a relevant sense right for me. However, this would be the very same sense in which we could say that the cat should not eat the whole stick of butter, because it will get sick later. Although we might be able to sensibly interpret what the cat should do from its rough collection of interests, this is not a sense in which the cat has normative reasons from the first-person, deliberative standpoint. If I do not have the special aim, according to constitutivism, then I am not an agent; if I am not an agent, then first-person practical normativity does not exist for me. Normativity reaches a determinate boundary where the normative standards stemming from the structure of agency do not apply. The scope of first-person, practical normativity includes all and only rational agents. A constitutivist theory will specify a limitation on who is an agent, which is, thus, a limitation on the scope of first-person, practical normativity. So constitutivist theories provide insight into the limits of practical normativity, which are otherwise rather vague.

A second corollary concerns the connection between deliberative, practical normativity and third-personal normative assessments. A consequence of a particular constitutivist theory is a demarcation of the applicability of normative assessment. Once again, there are senses in which we can say that the cat should not eat the stick of butter, or someone who does not have the constitutive aim of agency should maintain good health, but these senses are not very interesting for the investigation of normative reasons. If we focus our attention on normative assessments which purport to track practical reasons, then normative assessments like "what he did was wrong" or "it would be right for her to do that" are tied to the same limitations as first-personal normative reasons, because they are also tied to the agency of the acting party. That is, when the boundary of agency set by a particular constitutivist theory is crossed, third-person, normative assessments are not appropriate, at least strictly speaking. Since third-person normative assessments are strictly only applicable where they bear some relation to the first-personal normative standards of the assessed agents and their actions, third-person normative assessments do not apply to those who do not pursue the constitutive aim of agency. So, if I do not pursue the constitutive aim of agency, then it makes no sense for you to assess my behavior as right or wrong in the full, practically normative sense.

The relation between third-personal and first-personal reasons may be very minimal, as it is assumed to be on consequentialist normative theories. Consequentialist theories hold that action is right or wrong in accordance with some specified consequences that follow from it, whatever first-personal reasons the agent of the action may have had. However, it is still required that the agent has first-personal reasons of some kind or other. If an individual does not operate on deliberative normative reasons,

then third-person assessments of the individual's behavior are inapplicable, even on a consequentialist view. Although we may sometimes say that the dog did something bad, or that the cat did the right thing, and such behavior may even have some proto-moral character, these judgments, when made in an unqualified way, are either misplaced or insincere. Similarly, any individual that is not an agent, and so does not act on normative reasons, is not subject to normative assessments of the kind that are to be explained by constitutivism. So, if S does not pursue the constitutive aim of agency, then S is not an agent who performed an action, and so a normative assessment of S's behavior is not applicable. I defend and elucidate this point further in Sections III and VI.D, in my discussion of the Assessability Thesis.

B. Enoch's objections

Enoch's arguments are meant to undermine the constitutive approach to normativity, which I am employing in the explanation of practical normativity I offer. After a brief review of Enoch's objections, I defend the constitutive approach against Enoch's attacks and attempt to show that Enoch simply demands what should not be demanded of an explanation of practical normativity, and he fails to appreciate how constitutivism is supposed to work. I try to argue that although Enoch says he is granting the constitutivist analysis, he does not really take it seriously. If we did have an analysis of agency in terms of a constitutive aim then some of Enoch's crucial claims would be faulty. Enoch starts by pointing to a feature of the constitutive approach to normativity, which is absolutely fundamental to it – that practical normativity springs from the constitutive aims of agency – and he argues that, as a consequence, there is no normative justification for following the constitutive aims of agency themselves. This is absolutely

correct, but Enoch goes on to argue that this fact undermines the potential for explaining practical normativity in constitutivist terms, which I will attempt to refute.

The problem, according to Enoch, is that the normativity of what follows from the constitutive aim is only as good as the normativity of that aim itself, and it cannot have normative force (on a constitutivist system), because it is supposed to supply the explanation of normative force. If I am supposed to have a reason to eat better that comes from an aim to be healthier, then the normativity of my eating better is only as good as the normativity of my aim to be healthier. On a constitutivist system, however, the constitutive aim cannot have normative force itself, so if the normativity of whatever standards follow from the constitutive aim depends on the normativity of the aim, then constitutivism is in big trouble.

One way that Enoch presses this point is by asking us to consider an alternative to following the constitutive aim. I might not follow the aim and so, according to constitutivist thinking, I would not be an agent. Instead, Enoch suggests, I might be a "schmagent" – someone as close as possible to an agent, but who does not subscribe to the constitutive aim of agency. If it is suggested that I, a mere schmagent, should conform to some normative practical standard following from the aim, I can shrug this off with impunity, because I do not strive to be an *agent*, and I do not pursue the aim of agency, and so I have no reasons relative to the aim of agency. This would mean, however, that I am not subject to reasons at all, according to the hypothesized constitutivist theory, because all practically normative reasons are supposed to be reasons relative to the constitutive aim. The whole point of constitutivism, however, is to explain how it is that I *am* subject to reasons.

Another way that Enoch presses the point about the inability of the constitutive aim to supply its own normativity is by turning the examples used by the constitutivists against them. Korsgaard argues that one always has reasons to build a house according to empirical standards of good house construction, because if one is not trying to build a well-working shelter for comfortable living, etc., one is not trying to build a house. So, if anyone denies that they have reasons to build a well-working shelter when building a house, they can be refuted immediately by pointing out that a well-working shelter is part of what a house is and so they have reasons to build a well-working shelter insofar as they are building a house. The constitutive standard does the work of supplying the reasons. Velleman argues in a similar spirit that, when playing chess, one always has reasons to make the most effective moves toward an eventual checkmate, because if one is not, at least in some sense, trying to checkmate one's opponent one is not really playing chess. It of course seems wrong that one cannot play chess badly, or half-heartedly, because to do so would not really be playing chess. However, Velleman allows for this and points out that someone playing chess, "may even have the goal of losing a game . . . But in order to lose a game of chess, he must stay in the game, by continuing to pursue its object, however insincerely or ineffectually" (Velleman 1996, 713). So, according to Velleman, insofar as one is playing chess, one has reasons to work toward an eventual checkmate.

These examples do a nice job of displaying the work that constitutive standards can do within a specific domain, but they of course only represent cases of having reasons relative to some aim, where the aim itself has no special status except that it is constitutively attached to a functional concept like 'house' or 'chess'. Enoch uses the

relative arbitrariness of the corresponding aim and concept to turn these examples around and make his own point. Truly having reasons stemming from constitutive aims, it seems, requires having reasons to follow or have the aims: I only have a reason to build a house according to empirical standards of good shelter-building to the extent that I have a reason to build a house, and I only have a reason to move in chess in accordance with the object of pursuing checkmate to the extent that I have a reason to play chess. If I have no reason to build a house or to play chess, then I have no reason to build a good shelter or pursue checkmate – at least, no reason stemming from the constitutive aims of house-building or chess. If you and I are moving chess pieces according to the rules and I am only haphazardly moving them around, then whether you call what I am doing "playing chess" puts no additional pressure on me to try to work to eventual checkmate. If I have no reason to play chess in the (apparently) proper sense that involves a genuine effort to reach checkmate, then I really have no reason to make any particular moves that would facilitate doing so, even if I jeopardize my behavior counting as "playing chess."

According to Enoch, since the normativity of derivative standards must come from the *normativity of* the constitutive standards if it comes from the constitutive standards at all, the general explanation of normativity that the constitutive standards are supposed to supply would end up being circular. The constitutive aim cannot supply an explanation of its own normativity in any desired sense. So, one cannot have a normative reason to follow the constitutive aim, and having such a reason is, it seems, needed to get general practical normativity up and running via the constitutive strategy.

This is a core issue for a potentially profound way of dealing with metanormative questions. Enoch has touched upon something that may not be as detrimental to

constitutivism as he thinks (or even damaging at all), but which is a fundamental point to be brought out and examined for the constitutivist program. What is needed to address this issue is to lay out Enoch's claims and compare them with what constitutivism implies. Doing so reveals that what Enoch is pointing to is not any incoherence or weakness in the constitutivist approach but rather an acceptable and very interesting consequence of the view. This examination shows that what is really at stake is drawing up the limits of practical normativity.

C. Refutation of Enoch

To return to the problem, the fact that one has no first-personal normative reason to follow the constitutive standards for action, agency or deliberation is irrelevant to whether an explanation of general first-personal normativity has been given. We are trying to give a third-person, theoretical explanation of first-person, practical normativity. It is true that constitutivism does not allow us to do this with regard to pursuing the constitutive aim – it does not allow us to explain why the constitutive aim itself is normative for an agent.

However, no such explanation is needed. Let us look again at Enoch's objections. I read Enoch as having two distinct arguments against constitutivism, both of which are important and demand careful examination, and both of which make use of the observation that constitutivism cannot explain any fundamental sense in which the constitutive aim itself is normative for an individual. They can be summarized in the following way.

Argument 1
Enoch's First Argument: Schmagency

i) If I don't follow the aim, this may make me a mere schmagent and not an agent, but the constitutivist must still explain why the derivative standards are normative for me. Why I have normative reasons for anything must still be explained.

ii) The "No aim implies no agency" claim is the constitutivist's only resource. The constitutivist cannot explain why I have a normative reason *to be an agent*.

iii) Therefore, the constitutive program fails to fulfill its assigned task – to explain practical normativity in general.

Argument 2
Enoch's Second Argument: The Game Analogy

1) The constitutive aim of agency can only transfer its own practical normative force to derivative standards. It cannot initiate normativity from scratch.

2) If the aim of agency itself is normative, that normativity cannot be explanatorily derived from the aim. Its normativity must be explained by something else.

3) Therefore, the constitutive aim of agency cannot provide an explanation of practical normativity.

In both cases I deny the first premise. One could really just accuse Enoch of begging the question on these points, since he assumes constitutivism cannot explain normativity by the role the constitutive aim plays in agency and practical thought. I do more than this by arguing that some of his premises are false and that constitutivism can explain practical normativity in the intended way. The refutation of the First and Second arguments

follows from the forward and backward directions, respectively, of the following biconditional, which presupposes that we have some established aim of agency.

(*Assuming a successful constitutivist analysis of agency*)

(CA) Practical normativity exists for an individual if and only if that individual pursues the constitutive aim of agency.

(CA), once substantiated, refutes both of Enoch's arguments and legitimate the constitutivist strategy. The forward direction, which I will state in its contrapositive form for clarity in the succeeding discussion is the following:

(CA_1) If an individual does *not* pursue the constitutive aim of agency, then practical normativity does *not* exist for that individual

As I have already argued, the first person, deliberative standpoint is the standpoint from which one can have practical or normative reasons. The first of the corollaries I argued for, is that the constitutive aim of agency, if there is one, puts a definitive limit on practical normativity for an individual in the first-person standpoint. That is, an individual not pursuing the aim implies not being an agent, and not being an agent implies that practical normativity does not exist for that individual; hence, practical normativity exists for an individual only if that individual pursues the constitutive aim of agency.

This argument for (CA_1) can be stated formally in the following way.

Argument 3

1. It follows from the assumed constitutivist analysis of agency that if one does not follow the constitutive aim of agency, then one is not an agent.
2. If one is not an agent, then one does not operate on practical normative reasons.

3. Therefore, if an individual does not pursue the constitutive aim of agency, then practical normativity does not exist for that individual.

This takes care of Enoch's First Argument. It follows from (CA_1) that premise (i) of the First Argument is false. The constitutivist does not have to explain normativity for schmagents. Herein lies Enoch's fundamental mistake. Enoch argues that there could be individuals who are as close as possible to agents, but who do not have the constitutive aim. For these individuals there would be no reasons that stem from the constitutive aim, and so constitutivism would be left unable to explain any sense in which they had reasons. Enoch concludes that constitutivism fails to explain all the kinds of reasons we need to explain, and so fails as an explanation of practical normativity. With this argument, Enoch is directly pressing constitutivism to explain first-personal normativity for individuals who do not have the constitutive aim, and so, by hypothesis, are not agents. But for those who are not agents, there is no practical normativity and so there is nothing for a constitutivist theory to explain. Enoch does not take the assumption of a successful constitutive analysis seriously enough. He demands that an explanation of normativity be given for those who are not agents.

I should point out that schmagents, those as close as possible to agents but without the aim, might have the ability to trade in normative semantics. For all constitutivism of the practical variety we are discussing has to offer, there may be individuals who have sufficient mental structure to think theoretically in a relevant way, but who are not rational agents. Such individuals would not act on reasons, and so, even if they could consider questions like "Why should I be moral?" and understand what they

were saying, they would have no reasons to be moral, or to not be moral or to do anything at all, in a practical normative sense. It may seem strange to separate the linguistic ability to trade in normative semantics from the structural practical characteristics of agency, but at the most basic level, the linguistic ability is conceptually distinct from being a rational agent. Even if there is an underlying necessary interplay between theoretical and practical capacities, this is not obvious. So, the ability to ask "Why should I follow the constitutive aim of agency?" does not obviously imply that the questioner must be included in the group of all those who have practical reasons that need to be explained. Whether or not I *should* do anything is a matter of whether or not I am a practical agent. If mere schmagents have the ability to trade in theoretical normative assessments and to ask normative questions, this just shows that the category of those with this theoretical ability must be distinguished from the category of those who have practical reasons to explain. It follows from a particular constitutivist theory that first-personal practical normativity is limited by conformity with the constitutive aim, and so this aim sets the standard for who has normative reasons at all.

That there are such strict limits to practical normativity that are directly tied to the pursuit of some aim which is quite contingent may seem like a radical idea. However, as stated in the second of the corollaries I argued for above, for those who do not pursue the aim, the behaviors of those individuals are not normatively assessable from a theoretical standpoint, because they are not the actions of agents, so, there is nothing wrong with the result that there is no explanation of the normativity of any aims or inclinations those individuals might have – nothing is practically normative for such individuals. Individuals who do not pursue the constitutive aim of agency are not agents, and are

certainly not agents performing actions. So, there are no normative standards for such individuals. There is nothing to explain.

The backward direction of (CA) does the work in refuting Enoch's Second Argument:

(CA_2) If an individual pursues the constitutive aim of agency, then practical normativity exists for that individual

This would show that premise (1) of the Second Argument is false and would directly validate the explanatory potential of a constitutivist theory. If this can be verified, then constitutivism can explain practical normativity for all agents (at least in some minimalist form). In what follows, I present an argument for (CA_2) assuming that the philosophy of action aspect of constitutivism can be satisfactorily carried out in the appropriate way. To facilitate this argument and use the assumption that a constitutivist analysis of agency can be made out, let us call the thing aimed at according to some presupposed constitutivist analysis the "constitutive element." So, if the constitutive aim of agency is an aim at the unity of the agent, then the unity of the agent is the constitutive element.

For those who do pursue the aim, given a particular constitutivist theory, an explanation of general normative standards for that agent is readily available. Specifically, it can be explained why the agent should do that which helps achieve the aim. To start, the agent has the aim and there are things which are conducive to pursuing it, so the agent has reasons relative to the aim. But, the aim which such reasons are relative to marks success or failure as an agent. Since the aim is constitutive of agency, it is what one pursues as an agent. One's agency, or in the particular case, one's deliberation, is aimed at the constitutive element. So, to succeed as an agent, or to

succeed in deliberation is to attain or promote the constitutive element. Since the constitutive aim marks success as an agent, and the standard of success as an agent is the most general practical standard, the reasons must be reasons *simpliciter*.

Another way to see the same point is to think in terms of governance. Practical normativity, or practical reasons, is the regulative element in governing choice and action. If we can show that there is a regulative element which governs choice and action, then we will be able to simultaneously show that there is practical normativity – genuinely practical reasons. Now, the aim of agency governs choice or action in the sense that it sets the standard for successful or correct choice or action. This is because that which is aimed at in making choices or performing actions is the constitutive element. Just as the standard for successful bread-making comes from what a person is trying to do when they make bread, the standard for successful choice or action all together comes from what a person is trying to do when they make a choice or perform an action. That is, the standard of success stems, not from what one is trying to do in the particular case with the particular action, but rather, it comes from what one is trying to do in performing an action at all. When I deliberate about what to do, I do so successfully just in case I do what my deliberation is aimed at, in deliberating. Insofar as I make a practical choice, I am, in making the choice, aiming at the constitutive element. Thus, the constitutive aim of agency governs choice and action. It provides a standard for successful choice and action, or, thought of more generally, successful agency. Since practical normativity tracks success as an agent, or success in choice and action, a standard of success as an agent is a standard of practical normativity. Therefore, if an

individual pursues the constitutive aim of agency, then practical normativity exists for that individual.

The argument for (CA_2) can be stated formally as follows.

Argument 4

1. Assume that there is a constitutive aim of agency and that an individual has this aim in the appropriate sense.
2. Since the aim is constitutive of agency, it is what the individual pursues *as an agent*.
3. So, to succeed as an agent is to do what is targeted by the constitutive aim.
4. Practical normativity tracks success as an agent.
5. So, a standard of success as an agent is a standard of practical normativity.
6. So, the individual has a standard of normativity.
7. Therefore, if an individual pursues the constitutive aim of agency, then practical normativity exists for that individual.

One might object that I have not ruled out the possibility that when an agent deliberates the agent does so successfully if they get it right by some external, aim-independent standard. In such a case, it would seem that the constitutive aim of agency did not govern choice and action, since it would not be the standard of successful choice and agency. However, one must remember that, at the moment, what I am arguing is that if one pursues the aim of agency, then practical normativity exists for that individual. In the argument I assumed that there is no external, aim-independent standard of successful deliberation, because if there were then practical normativity would exist for the individual. So, if an individual pursues the aim of agency then either there is an external,

aim-independent standard or there is not. In the case where there is, there is practical normativity for the individual. In the case where there is not, there is practical normativity for the individual, because the constitutive aim governs the individual's choice and action and thereby sets up a standard of practical normativity.

Thus, a successful constitutivist theory can explain practical normativity. Assuming a successful constitutivist analysis of agency, *if* an individual pursues the constitutive aim of agency, *then* practical normativity exists for that individual. So, contrary to Enoch's claims, the constitutive aim does not have to be normative itself in order for its derivative standards to be normative for an agent. Normativity really can be explained from a constitutive standard. What Enoch misses is that it is not the *normativity* of the constitutive aim, but its role in the structure of agency and practical thought that gives rise to the normativity of derivative standards.

Hence, the biconditional (CA) holds: practical normativity exists for an individual *if and only if* that individual pursues the constitutive aim of agency. This means that a constitutive aim can explain practical normativity, and that the first premise in each of Enoch's two arguments is false. Arguing for (CA) simultaneously refutes Enoch's objections, which are natural worries for constitutivism, and legitimates constitutivism as a potential way of explaining practical normativity.

One might still object that normativity for all agents is explained only at the cost of constricting who counts as an agent, but this is to bring the issue back to the philosophy of action and the particular arguments about what is constitutive of agency. That, of course, is a different matter, and is exactly the matter I pursue in my arguments for my particular constitutivist theory.

III. Actions performed by agents

To lay the groundwork for the arguments I make in this chapter and the next, I want to start by examining the core concepts I draw upon, how they relate to each other, and to summarize the arguments that set up my account of practical rationality. In some ways, the most basic term to look at is ‘action’. I use the term ‘action’ to mean an action performed by an agent, which implies that the agent has ownership over the action, and that the behavior which instantiates the action is under the agent’s control. So, whenever I use the term ‘action’ I take this to mean that it is a performance owned by an agent. There are, however, other senses of ‘action’ that relate to human agency such as ‘intentional action’. ‘Intentional action’ is not equivalent to ‘action performed by an agent’. Intentional actions, as I understand them, only need to be caused by a desire and corresponding means-end belief. If I have a desire (perhaps sudden and fleeting) to knock over an annoying and obnoxious person’s cup of coffee and a belief that hitting the cup with my hand will knock the coffee over, which jointly cause my hitting the cup over, then this is intentional action, in the relevant sense, although it is probably not an action performed by an agent (me), because there is nothing in the desire or the combination that makes the behavior attributable to me, the deliberating agent. This is assuming that this is peculiar behavior for me, caused by a momentary desire that has little to do with my character. The behavior is not, in such a case, in any meaningful sense, the result of a decision of mine.

The important point is that in order for an action to have normative significance – in order for it to be apt for rightness or wrongness – it must belong to an agent in a sense

strong enough to capture the notion that it is the result of some decision *of the agent*. I do not take the typical philosophical or ordinary language understanding of 'intentional action' to meet this requirement, and so I do not make use of or attempt to analyze this term. Instead, I focus on 'action performed by an agent', and look to see what is necessary for this to come about. I look for what it takes for an agent to produce and own something which qualifies as an action performed by that agent. There are several layers to this investigation, which overlap with different parts of my constitutivist explanation of practical normativity. I argue for a conception of *agent-causation* – the causation or control of behavior by an agent which constitutes the ownership of action by that agent. Thereafter, I extend this argument to an argument for a constitutive aim of agency, which actually helps fill in the conception of agent-causation. These two components, each of which comes in multiple stages or layers, jointly form a view on what it takes for an agent to perform an action as the owner of that action. This is one of the foundations around which my theory of practical normativity is built. The other way of looking at my theory is in terms of constitutivism, according to which the theory is geared at uncovering a constitutive aim of agency and supporting it with other theses that together jointly imply a normative conclusion.

To give a first look at my analysis of actions performed by agents, let me summarize what I argue is necessary and what is involved in an action performed by an agent. Some of this material is expounded and argued for in the Agent-Causation section, while some of it is handled in the Unity section and some in the next chapter. The first, and perhaps most basic point to note, is that at the most abstract level agency is unified authorship from the complete motivational field of an individual. Agency must be based

on the complete motivational field of some individual (which in principle could be a group of people or some other interesting collective), because that is what individuates the agent – it is what makes it the agency of some particular individual. More importantly, agency is unified authorship of that motivational field. The authorship, control, or decision must be unified in the sense that one's field of motives must be brought together to form a single decision that comes from the agent's decision-making base – one's field of motivation as a whole. In order to uphold a distinction between a decision that belongs to the agent and for which the agent is responsible, and the mere execution of an impulse that does not or does not fully belong to the agent, we need the concept of unified authorship. Similar to the previous point about intentional action, but taken in general terms, sometimes a person does something that we say is not really attributable to that person. This could be behavior like mindlessly swatting at a fly, saying something rude under stress that is surprising to the speaker, or failing to perform to the level of one's training on an exam. If I did such things, I might rightly say, "that wasn't me." For such behavior, we normally maintain a category of acts that are not attributable to an agent, and in particular they are not attributable to the individual who acted out the behavior, so to speak. This category is made possible by thinking of agency in terms of unified authorship – authorship or control that draws on one's motivational field as a whole to form a unified decision. This, I argue at length throughout this chapter and the next, requires some kind of unification of the motivational field. It is this bringing together of motives that makes a decision belong to an agent rather than just following from the force of some particular motive or group of motives.

The larger question, however, is how, in conceptual terms, this can happen. We need to know what is, must or could be involved in unified authorship from the motivational field of an individual. One way to frame this is to look at what the "role of the agent" is, as Velleman puts it (Velleman 1992, 461-481). That is, we want to look at what it is that we expect the agent to do in making a decision, or how it is that the agent makes the decision belong to it. The answer, as I argue following Velleman, is that the role of agent, what the agent does to control behavior, is to intervene between motives and intentions and to select from among them. It is the intervention which is key. In terms more familiar to Korsgaard, the agent endorses motives through reflection. Through reflection, there must be a process that intervenes between various motives one has anyway to endorse or select from among them to come to a decision. This is the opening for agency to enter into decision making and make way for the control of behavior or ownership of action. This is the place where the activity of the agent should be located in the process of behavior being caused by motivation and representation of the world. This is another layer to the argument.

But now, reflective intervention and endorsement require a certain structure. In order for it to be the case that a reflective process intervenes between motives and endorses and selects in such a way so as to embody unified authorship from the complete motivational field, there must be certain structural features to the intervention and the resulting decision. The actual intervention and endorsement must take the form of putting together a principle of action that constitutes the decision. The intervention must be manifest in a principle of action, because the principle is what captures the unified perspective of the agent in that decision. It is what makes the intervention the

intervention of the agent rather than just another impulse. The principle is what marks the decision of the agent. Now, since many or most of our acts do not directly result from forming a principle of action, many or most of our acts will not be actions attributable to us as agents in the full sense. However, they may still very well be attributable to us in a derivative sense, in which they result from training or habituation that we do on the basis of principles of action. So, there is a very important, broader category of acts attributable to agents that are not actions performed by the agents in the full and direct sense. We may not usually think of cooking eggs as only indirectly attributable to the cook because there was no immediate principle of action, but if it is true that the behaviors involved in cooking the eggs do not result from any reflective selection of what to do, then there is no other way to treat the matter. The agent's only way into the ownership of an act is either through reflective deliberation that results in a principle of action, or through the construction of character that leads to the behavior. In the case of the character being responsible for the attributability of an act, that character must in turn have resulted from preceding reflective deliberation or more character that somehow can trace an ancestry of attributability back to reflective deliberation. All attributability must ultimately stem from reflective deliberation, according to my arguments.

The other crucial aspect to how reflective intervention can wind up being unified authorship from one's motivational field is the aim of agency. As I said above, there must be a reflective process that does the work of intervening, endorsing and forming a principle of action. This process itself, however, must have its own unity or autonomy in order for it to do what it is supposed to do, in order for it to perform the role of the agent and to amount to unified authorship of the motivational field. There must be something

that makes it the particular process that it is, rather than just the piling on of more motives. There must be something that makes it a process that comes to endorse in a way that represents the whole motivational field. What gives the unity, the autonomy to the reflective process which is supposed to be the activity of the agent is a guiding aim. A guiding aim makes it the particular process that it is and allows that process to be identified with the activity of the agent. It cannot be just any aim though, since it must be something that allows the process to represent the whole motivational field in its endorsement and forming of a principle. The aim, I argue, is systematization. I first argue, in the Unity section, that the aim of deliberation, this very guiding aim of the activity of the agent, is an aim at the unity of the agent. This is just the simplest, most obvious way to think of what the aim must be, considering the work it is supposed to do. In the next chapter, I argue for a more refined understanding of the guiding aim in terms of systematization. What is systematized are basically schemes of principles of action. That is, we systematize our practical commitments – our formulations of our motivations in terms we use to organize them. So, the way we organize our motivations is of vital importance to how we, as agents, make rational decisions.

So, as I argue at length through this chapter and the next, the constitutive aim of agency is the aim at systematization of one's commitments to form a principle of action. That is what reflective endorsement does as a process that plays the role of the agent, and that is the activity of the agent that controls behavior. This layered understanding of an action performed by an agent gives us both a sense of what practical reason does in the individual in order to make normative concepts relevant, and provides us with a

constitutive aim of action which I use to explain practical normativity and the actual place for normative concepts with respect to our acts and actions.

IV. Initial presentation of the formal argument for a constitutivist explanation of practical normativity

Now that I have cleared up my general constitutivist strategy and cleared up the conceptual approach to practical rationality I employ, I present my specific explanation of practical normativity which employs this strategy. In this section, I present and describe my formal argument for my theory of practical normativity. First, I list and describe all of my premises for the version of the argument I put forward in this chapter. Only one of these premises is changed in the next chapter, the Unity Thesis. All of the other premises are part of my official theory and form the framework in which my Systematization and Deliberation Theses, which I argue for in Chapter 3, do their work.

Although my theory does, in the end, leave practical reasons contingent and non-universal in a deep way, most of the framework of my theory can be considered Kantian. As I said previously, constitutivism is a Kantian strategy, and I argue for a rather Kantian version of constitutivism. This Kantianism comes through in my insistence on the key role of principles, or maxims, and in my focus on the role of deliberation in the authorship of action, which is what Kant meant by his positive conception of freedom (Kant 1785,1911; Kant 1788,1913). Furthermore, one of my premises explicitly places rightness of action in the rightness of the antecedent deliberation, which is a thoroughly, if not exclusively, Kantian take on practical nature.

To further prepare the way to making my explanatory arguments, I now list my formal premises with descriptions of their meaning and significance. My full arguments for these premises come after I have formally stated the overall explanatory argument for my theory of practical normativity. So, first I describe the premises, then I state the argument, and then in the rest of the chapter I argue for the premises.

Agent-Causation: *For an agent to be the author of an action is for the agent to act on a principle which derives from that agent's reflective endorsement.* This is a statement of what is fundamentally required for an agent to cause behavior. Reflective endorsement is a concept I am borrowing from Korsgaard (Korsgaard and others 1996), who pinpoints the key to practical reason in the endorsement of ideas or motives by one's reflective activity. By 'principle' I mean roughly what is meant by 'maxim' in the Kantian literature, which is to say a universally formulated principle of action. As I discuss later on, this does not mean anything like a law of nature for all rational agents.

The primary thrust of this thesis is to express requirements for agent-causation, rather than to give an explanatory account of it. The requirements are what I draw upon in the argument for my theory of practical normativity – that is what the constitutivist strategy uses. However, I also present a sketch of an account of agent-causation based on the requirements and I argue that these conditions are sufficient for agent-causation and make it an empirically explainable phenomenon. This direction of the thesis serves to allay worries one might have about whether agent-causation and actions performed by agents are concepts that have no application. One might think that arguments showing the need for reflective endorsement to be a

process that intervenes between motives to form a principle of action, in order for the agent to perform and own an action, also show that an agent performing and owning an action could never happen because it could not occur in the empirical world. To answer this, I argue that a constitutive aim of agency that constitutes reflective endorsement solves this problem and makes the activity of the agent explainable, while also making the ownership of action explainable in terms of this activity of the agent.

Unity Thesis: *For a principle to derive from an agent's reflective endorsement is for it to come from deliberation which aims at the unity of the agent.* This is an approximation of Korsgaard's claim that the aim of action or agency is the unity of the agent. The idea is that the agent, to make a decision which is his or her own, must draw on something which constitutes the agent's decision-making ability. In order to do that, the decision must be made by an attempt to tie together all of the motivational forces available to the agent. Otherwise, the motivational force that wins out and carries the decision is what causes the outcome, and not the agent as a whole or the agent's reflective endorsement as such.

I say much more about this argument in the section on the Unity Thesis, and it should be noted that this is the overall structure of the argument for the Deliberation and Systematization Theses, which I argue for in the next chapter, and which replaces the Unity Thesis in the main argument. Thus, the systematization version of the argument, my official version, can be viewed as a more specific and precise development of the unity version. The systematization version of the argument also enjoys having clearer and sharper consequences, I believe, and so serves us better as a

theory of practical normativity. My arguments for the Deliberation and Systematization Theses in Chapter 3, which replace the Unity Thesis, will be carried out at considerable length and in considerable detail. So, one should not be too distraught if the argument I gave above, and even the argument I give in the section on unity, seem too quick or hasty.

Deontic Kantianism: *Right action follows from right deliberation, wrong action follows from wrong deliberation.* This thesis is not entirely radical, even though it may not be universally accepted. It is important to point out, however, that this claim follows rather easily from a more robust conception of what action is and from what is at stake in calling something a 'right action.' If we think of action in the way that I argue we should, which is as behavior caused by a principle drawn from the agent's reflective endorsement, and so is under "conscious control" to use Velleman's terms (Velleman 1996, 694-726), and if we think of right action as doing this well, then the deliberation is clearly what is to be evaluated in evaluating the action. The action is behavior caused by a certain pattern of deliberation. Doing this well would mean doing the deliberation well. So, the rightness of action is dependent on the rightness of deliberation. Again, I give a more detailed argument later on.

Attribution: *An act is attributable to a given agent if and only if it is an action that agent is the author of or it is behavior caused by habituation, training or preemption by (an) action(s) that agent is the author of.* Since so much of my theory focuses on the authorship, or ownership of acts and actions, it is important to see when exactly behavior counts as *owned* by the agent. The Attribution Thesis asserts that this

happens exactly when the behavior is, or is the causal result of full, rationally deliberated actions by the agent.

Here, again, I am following Korsgaard in her terminology. A *mere act* is behavior in response to representation that is not a full action, because it is not done on a principle from the agent's reflective endorsement. Mere acts are things like "mindlessly" swatting a mosquito biting your arm, using reflexes to turn and catch something you didn't see coming until the last minute, or, often, catching a ball while playing catch (when you know the ball is coming and watch it come to you). So behavior that does not make up a full, rationally deliberated action, but is the causal result of such, is what I call a *mere act attributable to an agent*. If I train myself on how to play catch by, in part at least, using rational deliberation about what to do (like maybe just how to practice), then the result of the training will be acts attributable to me, at least to the extent that the training I gave myself was responsible for the behavior in question, the catching of the ball. I do not claim that it is always clear and obvious whether a person performed a full action or just a mere act attributable to the agent, or whether and to what extent an act is attributable to the agent.

Transitivity: *A mere act attributable to a given agent is right or wrong in accordance with the rightness or wrongness of the action(s) from which it derives its attributability.* This is a simple point, but it must be explicitly stated for the argument to be formally correct. Since rightness follows deliberation, and attributability follows deliberation, rightness and attributability are bound together. It would make no sense for rightness to be tied to the deliberation behind action, in the case of full

actions, but to stray from this and vary with something else in the case of mere acts, which derive their attributability from the deliberated actions behind them.

Assessability: *An act is normatively assessable if and only if it is an act attributable to some agent(s).* This is a key claim for constitutivism in general. I have already argued for this above, in Section II. The idea is simply that agency, and actions by agents define the limits of practical normativity. That which is not an act attributable to an agent is not normatively assessable in any strict, theoretical sense. We may have practical reasons for holding people responsible for acts or behaviors which are not strictly attributable to any agents, or to the agents we hold responsible, but that is a practical, and not a theoretical issue. Theoretically speaking, normative assessability depends on attributability to agents.

Let us now look at the formal argument in full.

Argument 5

1) **Agent-causation**: For an agent to be the author of an action is for the agent to act on a principle which derives from that agent's reflective endorsement.
2) **Unity Thesis**: For a principle to derive from an agent's reflective endorsement is for it to come from deliberation which aims at the unity of the agent. (To be replaced in Chapter 3)
3) Thus, if an agent authors an action, then that action derives from deliberation aimed at unity. (From (1) and (2))
4) **Deontic Kantianism**: Right action follows from right deliberation, wrong action follows from wrong deliberation.

5) **Attribution**: An act is attributable to a given agent if and only if it is an action that agent is the author of *or* it is behavior caused by habituation, training or preemption by (an) action(s) that agent is the author of.

6) **Transitivity**: A mere act attributable to a given agent is right or wrong in accordance with the rightness or wrongness of the action(s) from which it derives its attributability.

7) Therefore, if an act is attributable to a given agent, then it is right or wrong in accordance with its compliance with the aim at unity. (From (4), (5), (6) and (7))

8) **Assessability**: An act is normatively assessable if and only if it is an act attributable to some agent(s).

9) Therefore, if an act is normatively assessable, then it is right or wrong in accordance with its compliance with the aim at unity. (From (8) and (9))

V. Agent-Causation

In this section, I argue for the Agent-Causation Thesis. My argument for the Agent-Causation Thesis will be carried out in four stages. First, I explain what I mean by 'reflective endorsement' and what role this concept plays in the picture of agency I am advocating. Second, I argue that action by an agent requires reflective endorsement. Third, I introduce another fundamental concept for my theory, that of a principle of action. Principles, in the sense I am using, are roughly the same thing as, or are an interpretation of Kantian maxims, as I explain. After discussion of principles and their role in decision-making, I argue that action by an agent requires determination of

behavior by a principle. Fourth, I sketch an explanation of agent-causation, the phenomenon of an agent causing behavior. While this does not amount to a full-blown account of the phenomenon, it goes some way in outlining one, and constitutes an argument that acting on a principle deriving from reflective endorsement is sufficient for agent-causation. Thereafter, I present a formal version of my argument for the Agent-Causation Thesis, which draws on the conclusions of the foregoing arguments.

A. Reflective endorsement

A concept I make a lot of use of in this chapter is 'reflective endorsement'. I am following Korsgaard in her use of the term 'reflective endorsement' (Korsgaard and others 1996), although I do not elaborate on it in quite the same way that she does. Reflective endorsement serves as the basic building block in my account of practical rationality. I use several arguments in this chapter and the next to show what reflective endorsement must consist in and what role it must play in practical rationality and practical decisions made by agents. Though a corresponding account of the role of reflective endorsement or reflective selection in theoretical reason could probably be given, my focus is solely on the functioning of practical reason.

To understand reflective endorsement, we must first understand the capacity which it employs: *Reflective consciousness* is the capacity to reflect on one's own mental states (beliefs and desires) and apply some procedure to select from among them those which will be engaged. Reflective consciousness operates by the application of the procedure (whatever it is) to other mental states, which makes those other mental states, according to David Rosenthal, conscious (Rosenthal 1986, 329-359). The selection which occurs through this reflective procedure is *reflective endorsement*. The procedure

of reflective selection, I later argue, must have some goal to make it a unified, singular procedure, but for now, we can treat it as a place-holder. (In Chapter 3, I argue that it aims at systematization.) Reflective endorsement, for now, can be thought of as whatever it is that makes practical selections through reflection of options by a survey of desires and beliefs. It is the reflective procedure that organizes and selects the beliefs and desires to be engaged practically, and so acted on. If I am deliberating on where to go to lunch, I will have a large collection of various beliefs and desires which bear on where I might go. My reflective consciousness gives me the ability to reflect on my situation and my desires and use a reflective procedure to select and endorse a particular option. Ultimately, I argue that the reflective procedure I must use to select from my options and endorse my choice, in order for it really to be the choice of me, the agent, is to work to systematize my field of commitments. This, however, will first require a lot of preliminary argument and development of concepts.

In the next section I argue that activity via reflective endorsement can be equated with behavior under conscious control, which is the trademark of action by an agent, and so reflective endorsement can be equated with the activity of the agent. This also accords with another argument I use later on, that the role of the agent is to mediate between one's various beliefs and desires to form intentions and to lead to behavior.[6] Mediating between a field of beliefs and desires is exactly what reflective endorsement does, which means that it plays the role of the agent in this sense. That which plays the role of the agent can, of course, be equated with the activity of the agent.

[6] This argument is taken from J. David Velleman, "What Happens when Someone Acts?" *Mind: A Quarterly Review of Philosophy* 101, no. 403 (July, 1992), , 463.

B. Conscious control

In this subsection, I aim to draw a definitive link between what it takes for an agent to perform an action and reflective endorsement. I aim to show that there is a certain, identifiable structure to action by agents, which has some very interesting consequences down the road. I argue, namely, that *in order for an agent to cause behavior and perform an action, the behavior must be determined by the agent's reflective endorsement*. In order to make this argument, I first give a general account of action as behavior under conscious control which I am borrowing from Velleman. I begin here, because I believe that this analysis of action is fairly straightforward and is easily accepted. I argue that it should be accepted without reservation as a basic analysis of the notion of 'action by an agent' as this notion is relevant for ethics. Then I present a reflective endorsement account of action, which has more complexity than the initial conscious control account, and I argue that the two are equivalent. Again, I take the route through the conscious control analysis just as a means to make the argument for reflective endorsement.

An *action*, according to Velleman, is "behavior executed under conscious control" (Velleman 1996, 718). This definition is clearly a correct understanding of action in the context of analyzing the nature of our normative experience for the following reasons. Obviously, action involves behavior of some kind. But there must be something more to it, something which differentiates it from just any behavior which might be uncontrollable or might be something an insect could do. Humans have a special power, or so we think, which allows us to perform actions in some sense which makes us special and which makes us susceptible to normative demands. Velleman pins the special power on a supposed capacity for conscious control. Behavior not under my conscious control

does not constitute my action, he argues, just as much as conscious insistence without behavior does not constitute my action.

Velleman is clearly right in his analysis. It takes little to see that conscious control is the component of action that works with the overt behavior to make it an action by an agent that is interesting for moral philosophy and normative evaluation. This argument is easiest to make if it is aimed directly at the normative evaluation result, we want to get to later on with the Assessability Thesis. In other words, I want to focus directly on 'action performed by an agent' in the sense which makes normative evaluation of the action applicable. Obviously, we are talking about behavior, but it must be under conscious control or it cannot be related to what one *should* do, what one is responsible for, or what it is good to do.[7] Conscious control is necessary for normative evaluation to be applicable. In a certain sense, conscious control is also sufficient (or perhaps quasi-sufficient) for behavior to be subject to normative evaluation, since only this definition, or one which is pragmatically equivalent to it or implied by it, is interesting for the purposes of metaethics. If some behavior of my body is under my conscious control, then it is subject to normative evaluation. The question of what I should do arises.

So, I take 'action', throughout my investigation, to mean behavior under conscious control. An action is the whole package of behavior and the conscious control it is under. Action is not equivalent to the behavior component. Thus, an agent does not

[7] One might be concerned about how subordinate processes fit in, like the non-conscious control I may have over the way my feet move when I walk, or continuing to sleep when I should have already gotten up. However, insofar as these are actions or there is an action involved in the situation, it is the conscious control that I have had or do have over the non-conscious process which makes this so. I may consciously allow myself to walk in my usual fashion, just as a manager allows employees to work as they always do, or I may occasionally consciously attempt to control my foot movement while I walk which is responsible for how I walk in general (this follows from the Attribution Thesis to be discussed later on). Similarly, the steps I take to wake up on time are consciously controlled insofar as they are actions by me.

cause an action; an agent *performs* an action. The behavior component is caused or controlled by the agent, which, all together, makes up the action the agent performs. Behavior is one component of action, while conscious control, in whatever form it does or must take, is the other component. I may perform an action of helping a friend, which already includes the conscious direction in the description. The behavior would just be all the activities involved in whatever is being done, like stopping by the store to pick up an item. The conscious control component is implied by the fact that I do this *as* helping out a friend. Through a reflective process, I cause my behavior that adds up to running the errand. Throughout this discussion I keep my use of 'behavior' separate from 'action'. I also sometimes use the term 'act' to denote something that is not necessarily a full action, but is understood as meaningful behavior that might well be attributable to the agent.

This general notion of action, behavior under conscious control, is equivalent to the reflective endorsement analysis in the following sense. One cannot reflectively endorse unconsciously. One can only reflectively endorse through consciousness – reflective consciousness is a kind of consciousness. Thus acting according to endorsement, or acting as determined by endorsement, is acting as determined through consciousness – it is conscious determination. If my endorsement swings me to acting, if it determines the way that I act, then the act is controlled by the consciousness through which this endorsement happens – my endorsement is a controlling consciousness. Therefore, behavior determined by reflective endorsement is behavior under conscious control, which is action.

More importantly, for my purposes, I will presently argue that reflective endorsement is *required* by the conscious-control notion of action. This argument goes as follows: If my behavior is under my conscious control, it must first present itself to me as an impulse or desire – there must be a motivational force of some kind. For me to control which impulses I act on, I must choose which of them will be executed. For me to choose which of my impulses will be executed, is for me to endorse my acting on some but not other impulses through reflection. Therefore, for me to consciously control my behavior is for me to reflectively endorse my acting on some but not other impulses to behavior.

What has been gained by this analysis is the following argument: In order for an agent to cause behavior and perform an action, the behavior must be under the agent's conscious control. In order for behavior to be under an agent's conscious control, it must be determined by the agent's reflective endorsement. Therefore, *in order for an agent to cause behavior and perform an action, the behavior must be determined by the agent's reflective endorsement*.

C. Principles

A fundamental concept in my system of practical normativity is that of a *principle*. I use the term 'practical principle' or just 'principle' to mean a universalized hypothetical directive: "I do X, on condition Y." Examples of practical principles could be, "I use a PC instead of a Mac, because that's what I'm used to," or, "I give money to the United Homeless Organization, when I haven't done so in a while," or, "I add red wine to the vegetables, when I'm making chili," or, "I add milk to my coffee, whenever I have coffee that isn't very good." All of these practical principles, or principles of action,

have the structure of avowing categorically that the agent does something on a given condition. In the first example, a ‘because’ is used to signal the condition, by which it is implied that being used to something of some relevant class is a condition on which the agent sticks with that thing. This ‘whenever’ structure of principles of action is their most basic trait and is what makes them universal in the sense I am using.

Principles play a key role in my arguments to come, especially in the arguments of Chapter 3 for the forthcoming Deliberation and Systematization Theses. It is significant that a preliminary understanding of acting for a reason can be had based on the idea of acting on a principle, in the sense I am here expounding on. If I act on the principle, “I buy good books, when priced at a bargain,” then I can also say that I am buying the book because it is a good book and has a bargain price.

Although principles in my sense are maxims in the Kantian sense, I do not use the term ‘maxim’ because of the confusing connotations that term carries from its role in Kantian ethics. I also choose not to use the term ‘practical policy’ in my arguments, except when used for emphasis, because ‘policy’ also has the misleading connotation of not applying to small decisions, and of always being explicit. Thus, I use the term ‘principle’.

To help elucidate what I mean by ‘universal,’ and to distinguish it from what Kant meant, I want to elaborate on the status of maxims and laws as used in Kant’s ethics.[8] It is very important that principles are universalized, but this only makes them maxims, not “universal laws of nature” for all rational beings (Kant 1785,1911, 421) or anything of that sort. Maxims are already universal in my sense, because they take the form “I do X

[8] For a discussion of different conceptions of maxims, as well as an argument against the conception I am advocating, which I will respond to in the Attribution section, see Talbot Brewer, "Maxims and Virtues," *Philosophical Review* 111, no. 4 (October 1, 2002), 539-572..

on condition Y (universally)." Kant made a mistake in his famous argument for the Law of Nature interpretation of the Formula of Universal Law, and all of his analogous arguments, by doubling up on universalization (Kant 1785,1911; Kant 1788,1913). Since maxims are already universal in the sense relevant for agency, what would it be to universalize them? – A universal law of action. This second application of universality, however, is unwarranted by the argument. Maxims are already in universal form, so when Kant argues for the necessity of universal form, the maxims already have it. There is no impetus from the argument to put maxims to the universalization test.

We can see that maxims must already be in the universal form I am talking about, because they must be formulated as a policy in order for the universalization test, famous from Kant's ethics, to make any sense. "Whenever I believe myself in need of money, I will borrow money and promise to pay it back, even though I know this will never happen," (Kant 1785,1911, 422) is how a maxim is supposed to run, and is ready to be generalized into a law of nature, governing the conduct of all rational beings. However, "I like money and I am taking this money off the counter" is not a proper maxim, and is not ready to be turned into a law of action. We should not be tempted to read the 'and' as a 'therefore', because this is a way of sneaking universality back in. The 'therefore' would only make sense if 'I like money' is understood to be some kind of universal policy of action. Taking the non-universal reading of the maxim, what would the proposed law be? Any law of action must be in the form of a policy which dictates what is done on what condition. "I like money and I take this money" governing the behavior of all rational agents makes no sense, because it is not in the proper universal form of a general policy. The alternative policy has universal form of the kind I am claiming for

what I call "practical principles." The point of this discussion is to distinguish my position from Kant's position, and to make clear that what I mean by 'universal' is not the same as what Kant meant, and that it does not imply any laws for rational agents. It does, however, assume a distinct formal structure, which will be very important later on.

I also want to emphasize that nothing I am saying about principles entails that they cannot be changed in a heartbeat or on the slightest whim by the agent. While they take universal form in their application, and so, in a sense, carry a general "commitment," this does not in any way imply that the agent cannot change them or is stuck with a principle, once applied. The force of practical normativity does not come, on my theory, from a commitment to universalization, or the consequences simply of making decisions in universal form. That was Kant's theory, but it is not the theory I am advocating.

The point of emphasizing the need for universal form comes from the way it sets up the constitutive aim of agency. The aim of agency is to unify itself in its decisions. This works with principles, because an agent's decisions need to have some formal structure in order for 'unifying oneself' to have any meaning. If my decision is to unify my attachment to my family and my attachment to my career, I do this by making these two commitments into a single decision-making fund, so to speak, insofar as I formulate a principle which determines their respective roles in the context, like deciding where to move. By connecting my commitments to a context with a principle like, "I move within driving distance of my extended family to take a job, when a job of such and such a nature presents itself, etc." (the actual principle would be very complicated), I aim to unify my commitments. What counts as better or worse unification is a question for Chapter 3, where I move from a discussion of unification to a discussion of

systematization, which allows for a clearer understanding of the relevant standards. I argue that a more precise and refined understanding of the aim of agency shows that what is really at stake is the aim of rational deliberation at systematizing available principles.

To work further toward these grander arguments, I now want to argue for the Kantian thesis that *if behavior is not determined by a principle, then it is not the action of an agent*. This functions as a premise in the argument I make later in this section which concludes in the Agent-Causation Thesis. This premise ties action to principle-based decisions and so fills in the rest of the content of the Agent-Causation Thesis. This is an important step in the argument for the reasons stated above concerning the role of principles in the aim of agency.

To make this argument I first set up a dichotomy between acting on a principle and the alternative, acting on "singular edicts" in the words of G.A. Cohen (Korsgaard and others 1996, 176). Then I present Velleman's argument against what he calls "the standard story of human action" (Velleman 1992, 461), which points to the significance of agency intervening in the deliberative process to direct the decision. Then I ask whether the suggestion by Cohen and Michael Bratman of the possibility of a singular edict account of an agent performing an action can capture intervention. I argue that it cannot and that intervention of the needed kind can only be achieved through the use of practical principles. I conclude that the alternative to acting on a principle is not a possible way for an agent to perform an action. Thus, an agent must act on a principle to perform an action.

The alternative to acting on principles is acting on "singular edicts," as Cohen puts it. A singular edict is supposed to issue from agency but without the structure of

principles and universal form. Presumably, if I decide to take the train, a singular edict behind this outcome would be of the form, “I want to take the train now,” which then results in my taking the train. The decision must result from a desire, but the desire may be supported by other desires including higher-order desires as is suggested by the following example from Michael Bratman’s review of *The Sources of Normativity*.

> Suppose I desire to A now. Suppose I endorse acting now on this desire, and that this endorsement is itself part of the explanation of my so acting. Suppose that I endorse both this endorsement and its role in my agency. And suppose that there is not, in fact, any other present endorsement of mine that is incompatible with these endorsements. My relevant endorsements could be specifically concerned with my A-ing now in this particular circumstance; they could be “singular edicts” without a commitment to a universal principle of action. Yet such a structure of endorsements seems to provide some traction for talk of an agent who is not simply a place where desires push and pull. It seems to provide resources for the project of constructing the “active self” (Bratman 1998, 707).

Bratman’s example is left very abstract, which helps to cloud the issue and make it seem less objectionable that singular edicts could result in the agent’s decision. It seems, however, that the example would have to be filled in in something like the following way: I desire to take the train now. I like that I have a desire to take the train. My liking my desire to take the train is a cause of my taking the train. I like that I like my desire, and I like that my liking of the desire is a cause of my taking the train. There is no other present liking of desires that is in conflict with my liking of my desire to take the train. This is how Bratman’s example goes. He claims that the endorsements can be

singular (which is how I have framed them), and that this structure can be enough for agency, and not a mere pushing and pulling of desires. It is important that the endorsing is filled in in terms of something along the lines of liking, because leaving it in terms of endorsing allows the ambiguity and lack of clarity to loom over the issue. The term 'endorse' by itself does not indicate that any kind of universal form is necessary, according to Bratman. This is okay, but we need to know what 'endorse' means in this context. To endorse in completely singular terms, as Bratman would have it, can only be to like, desire, or just not dislike. Adding on structure only leads us back to categorical attitudes (holding the attitude toward a whole category) and universal form of some kind.

To see what is wrong with this suggested singular-edict account, I want to compare it to the "standard story of human action," according to which reasons, understood as pairings of desires and means-end beliefs, cause intentions and intentions cause movements which constitute an *action* if the causal processes are "normal" (Velleman 1992, 461). Velleman takes issue with the standard account of action, because the agent, or role of the agent has been left out. The role of the agent is "to intervene between reasons and intention, and between intention and bodily movements, in each case guided by the one to produce the other" (Velleman 1992, 463). While the standard story may be perfectly capable of explaining behavior as resulting from intentions, it cannot explain an action performed by an agent as such, because the activity and contribution of the agent is left out. Say that Susan decides to buy a new chair, and let us assume that this is a case of an agent performing an action. The standard story could explain this incident in terms of Susan's desires and beliefs and her resulting behavior that amounts to her buying the chair. This may be a perfectly acceptable explanation as

far as it goes, but it would not be an explanation of her action as an agent, because it would not capture the sense in which she, the agent, made the decision and performed the action. Susan, the agent's contribution to the process would be left out.

What is left out of the standard story is the role of the agent. The role of the agent, the place where the agent enters into the practical process, is to intervene between occurrent mental states like desires and even intentions and to select and endorse from among them. There are processes by which reasons are sorted through to form an intention, and the intention is tested before it can form movements. These processes perform the role of the agent, and so they are the activity of the agent. They are what *I* do. Outcomes which are the result of these processes are attributable to me, while those which are not the result of these processes are not attributable to me.

As Velleman argues, the processes of sorting through reasons (and reflectively endorsing) cannot be reduced to a story of normal causation from beliefs and desires, through intention and on to movements. The role of the agent is to intervene between these states, which is something they cannot do for themselves. Although one may be able to give a causal explanation of *behavior* without appeal to the role of the agent, this will not be an explanation of the action of an agent, since it would leave out consideration of that which makes the behavior *action*.

To return to Bratman's singular-edict account, let us ask whether such an account can capture the crucial element of intervention that is essential to the activity of the agent being part of the picture. The answer is that when we look at the example it is plain to see that there is no room for the kind of intervention we are looking for. The fact that I like my liking of my desire to take the train, adds nothing special to the equation other

than that more of my motives are piled on top of the push toward taking the train. There is nothing there that can carry the weight of being the reflective intervention that sets up the decision as unified authorship from the motivational field. The piling on of more motivational states of the same unstructured variety does not help. If we are looking for attributability, we need more.

Intervention is the agent putting its stamp on a decision. Principles encapsulate that stamp of the agent – the agent's mark of endorsing or directing the decision. Principles capture the sense of unity the agent needs in order to make its mark – in order to make an endorsement or to intervene in a way that brings together motives without just being another particular motive, preference, or desire. The agent needs this kind of unity to intervene and to endorse, because otherwise there is nothing but individual motives. For there to be something over and above these individual motives, there must be a unity, something which is both formed from what materials are available, and which is a thing unto itself. Since there cannot just be another kind of thing that sits next to the motives and acts like a ghost in the machine, there must be a structure from what is already there – there must be unity of motives and attitudes that forms the intervention. This is agency and in particular this must be manifested in the decision as a principle.

So, Bratman and Cohen's singular edicts alternative leaves no room for the role of the agent, no room for reflective endorsement, despite Bratman's use of the term 'endorse'. However elaborate the web of desires may be from which the "edict" results, it remains only a pattern of activity of the motivations within the individual. It does not include a reflective process that intervenes and selects so as to reach a decision that amounts to unified authorship from the individual's motivational field. I do not choose to

A in Bratman's example, because for that to happen, I must intervene between and support the connection between my motives and my intentions, my intentions and my behavior. This operation cannot be performed by another desire of some higher-order alone, because that only introduces some new thing which I have to reflect on. There is no intervention, no point of entry for agency, unless there are principles of action appealed to. When the deliberation takes the character of, "I could do X, *because* Y," and, "I do Z, on the condition W," then we have intervention. Principles give a cohesive structure to deliberation and to decisions that is otherwise lacking. The process of reflection only ends in my judgment, when it ends with a declaration of what I do. I choose to do A when my reflection comes to a principle that I do A in the given situation, by whatever description I used in deciding whether to A. Therefore, *if behavior is not determined by a principle, then it is not the action of an agent*.

Velleman concludes from his argument that actions by agents must be done on reasons. This must, in the end, amount to the same point. When we speak of 'acting for a reason' we do not mean 'reason' in the sense of an objectively valid norm. Instead we mean to point out something about the form of the decision. We mean to say that a decision was made on the basis of what the agent *takes to be* a reason. What brings principles and reasons together is that reasons must have a universal form, which universally connects some condition to some act type. Reasons, in the present sense, and principles are the same thing in different formulations. Since my system requires the principles formulation, rather than the reasons formulation, this is the conclusion I draw. To help clear this up, let us look at how reasons can be formulated in terms of principles. Principles are practical propositions which assert that I do something on a given

condition. Reasons are the conditions in the principles. So, for me to have a reason (in the individualized sense Velleman appeals to) to take the train home is for me to have a principle that I take the train home in some situation and for that situation to be at hand. The situation in question is a reason for me to take the train home. In general, for me to make a decision which will issue in my action, I must decide on a principle or a reason. I must decide that I will take the train for a reason or it is not really my *action* to take the train.

D. The problem of agent-causation

Putting the conclusions of the preceding subsections together, I have thus far argued that *if* an agent causes behavior, *then* the behavior is determined by a principle from the agent's reflective endorsement. That is the conclusion I need for my main argument for my constitutivist account of practical normativity. If I argue in the other direction that *if* behavior is determined by a principle from reflective endorsement, *then* the agent causes the behavior; that would constitute an explanation of the phenomenon of agent-causation. As initiated by Harry Frankfurt (Frankfurt 1971, 5-20) there is considerable discussion in the literature over kinds of motivational forces which can be equated with agency. However, the suggestions which come out of this work, such as second- and higher-order desires, are inadequate for the task. As is shown by the "role of the agent" argument used above, the causality of the agent cannot be equated with simple belief-desire combinations, no matter how complex.[9] I have argued that agent-causation requires that the behavior be caused by the reflective endorsement of a principle which

[9] For an analysis of the literature on the subject and an assessment of the significance of the problem see Velleman, *What Happens when Someone Acts?*, 461-481.

determines the behavior. If this is also sufficient for agent-causation, it would, thus, present a solution to a significant philosophical problem.

While this argument would be significant for practical philosophy, I do not need it formally for my practical normativity argument. So, whether, and to what extent I need such an argument is a question of how serious one takes the problem of agent-causation to be. If one thinks that there is no such thing as agent-causation beyond belief-desire causation, and so my arguments only show that there cannot be anything like a true action performed by an agent, then this is a serious issue for my theory. If one is very comfortable with the idea of agent-causation and is sure there is such a thing, then this is not so serious, and my other arguments do all that they need to without an account of agent-causation. However, since it is available to me, I sketch an account of an agent causing behavior with the resources within my theory, so as to make my theory more complete.

In order to put together such an account, I need to lay out the elements of action by agents, accounting for how everything fits in its place, how the requirements I set forth are fulfilled and how action is psychologically explainable. That I make action psychologically explainable is a crucial requirement, because that is one of the two parts that have to be brought together: psychological explainability and the role of the agent. The agent must be incorporated into an explainable causal picture. That is what the problem is.

The picture is as follows. Reflective endorsement plays the role of the agent – to mediate between desires and beliefs to form intentions, which cause behavior. The role of intentions is played by principles, although this is not important, since intentions can

be effectively dropped from the account anyhow. So, we need a way for reflective endorsement to play the role of the agent in a psychologically explainable way, where this terminates in a principle that determines the behavior.

To properly give a philosophical account of this, I would need to fill in more details and use more of the theory I argue for later on, namely, the aim of agency. However, since I am not directly concerned with giving such an account, except to show that I could plausibly do so, I only sketch this in at this point. The aim of agency is what makes the phenomenon psychologically explainable. The individual really aims to do X (to unify its agency, or more specifically, to systematize its commitments), in a psychologically real sense, which causes the behavior. What makes this aim special, is that all rational decisions must go through it, because going through this aim is what constitutes a rational decision (in the sense of practical, not theoretical rationality). So, if my desire causes my behavior without being called upon and endorsed by the constitutive aim of agency, the aim to systematize, then my behavior is not the result of rational, or practical deliberation, and so is not an action by me. This is the constitutivist connection between psychologically explaining action and practical normativity.

What remains then, is to explain how the constitutive aim of agency amounts to reflective endorsement and leads to a principle of action. While this connection is really be made in, and is the main focus of Chapter 3, I can sketch how this goes. The constitutive aim of agency is the aim of rational deliberation to unify, or systematize the agent's motivational forces to create a principle of action. Reflective endorsement is unifying the agent's motivational forces in deliberation, or so I argue. Thus, while this may not be a full-blown philosophical theory of agent-causation, this certainly shows

how agent-causation can be accounted for with the resources my theory has available. It is enough, moreover, to conclude that *if behavior is determined by a principle from reflective endorsement, then the agent causes the behavior*.

E. Agent-Causation Argument

I am now in a position to put these arguments concerning action and agency together to conclude in the Agent-Causation Thesis. The argument is very simple at this point and only involves drawing from the arguments that have gone before. I include the conclusion of the last subsection for completeness, even though the biconditional conclusion it allows me to draw is not strictly necessary for my main practical normativity argument.

Argument 6

1) In order for an agent to cause behavior and perform an action, the behavior must be determined by the agent's reflective endorsement. (Conclusion of IV.B)

2) In order for an agent to cause behavior and perform an action, the behavior must be determined by a principle. (Conclusion of IV.C)

3) Therefore, in order for an agent to cause behavior and perform an action, the behavior must be determined by a principle which derives from the agent's reflective endorsement. (By composition)

4) If behavior is determined by a principle which derives from an agent's reflective endorsement, then that agent causes the behavior and performs an action. (Conclusion of IV.D)

5) Therefore, an agent causes behavior and performs an action if and only if the behavior is determined by a principle which derives from the agent's reflective endorsement.

For the sake of brevity and elegance, I will state the thesis as follows.

Agent-Causation: *For an agent to be the author of an action is for the agent to act on a principle which derives from that agent's reflective endorsement.*

VI. Unity Thesis

Once again, the idea of unity as the fundamental aim of practical rationality comes from Korsgaard (Korsgaard 2002; Korsgaard 1999, 1-29). However, I do not take my formulation of the claim from her and the way I elaborate on it and argue for it is largely independent of her work. Furthermore, this is only a temporary insertion into my overall explanation of practical normativity. In Chapter 3, I replace the Unity Thesis with the Deliberation and Systematization Theses for which the Unity Thesis is only a sketch. The same can be said for the respective arguments I use to establish these theses. This argument for the Unity Thesis is only a sketch, but it does provide the outlines of the more detailed and formal arguments for the more detailed and formal conclusions of Chapter 3. I offer this sketch now, because my arguments for the more detailed and formal theses of Chapter 3 are rather long and complicated. It might be a distraction from appreciating my arguments of this chapter if I offered no preview of what the claims of the next chapter will are and what role they play.

Reflective endorsement was defined as a deliberative process. It is a procedure which reflectively selects motives to be acted on, in such a way so as to constitute the activity of the agent. In my definition of reflective endorsement, I left it open what tied the procedure together which mediates between one's various motives to come to the

decision of the agent. The point of the Unity Thesis is to express (a sketch of) what that procedure is and what its defining aim is.

As previously stated, the role of the agent is to mediate between one's various motives to come to a rational decision. Reflective endorsement is the power of reflective consciousness one uses in order to perform this task. With the Unity Thesis, I am bringing out the necessity of aiming at unity in performing the role of the agent. Stephen Darwall has an argument concerning the unity of the agent, which can be repurposed to give us a starting point for the work needed. In the text, Darwall is arguing against a strongest-desire model of agency and deliberation:

> Were we but a bundle of such individual preferences we would have no way of coming to and expressing one mind on the question of what to do; for we would have no perspective other than that internal to each individual preference from which to order our different individual preferences, consider how to deal with conflicts between them, and decide what *we* prefer, on the whole, to do. (Darwall 1983, 103)

Darwall's point is that a mere collection of desires or preferences as the basis for our decision-making does not amount to the decision of one author. The author of an action must cause an activity from the perspective of that author as a whole and single unity. There is no such perspective, however, if decisions are merely the product of one's strongest desire. There must be something more than the strongest desire, something which constitutes the view of the agent. There must be a unity of the agent's will. What this means is just that the activity of the agent must aim to unify the agent's motivations.

The activity must be a process that aims at a kind of unity in the motivational field of the individual.

I have already argued for the place of reflective endorsement in agency – reflective endorsement is the activity of the agent. Put into this context, Darwall's argument shows the need for a unified perspective that reflective endorsement achieves. The simplest version of this, indeed a mere formalization of this, is that reflective endorsement must be a process which aims at unifying the available motives, which is to unify the agent. To come to a reflective decision which performs the task of choosing as an agent from one's motives, one needs to bring one's motives together into a unity so that one achieves a perspective of the whole, rather than the perspective internal to each motive, as Darwall might say. For us to talk of the agent as one thing that mediates between desires to come to a cohesive decision, then the mediation must be done through unification of the available motives, and the process of mediation must be tied together by an aim at this unification. Therefore, *for a principle to derive from an agent's reflective endorsement is for it to come from deliberation which aims at the unity of the agent*.

To understand this conclusion, and to appreciate the power of this argument, we can look at the meaning of aiming at the unity of the agent and unifying one's agency. Unifying one's agency is making one's motives a cohesive system of decision. I have already argued that decisions must be made in the form of principles. Doing this, deciding on a principle, is already a step toward unification. When one declares that one does something in a given situation, one gives cohesive structure to one's decision, and so gives cohesive structure to one's motives as they concern that situation.

The further step in unification through decision is to give cohesive structure to one's motives beyond that situation. This is the full meaning of aiming at unity or unification. If I decide whether to go see a movie or stay at my computer and work, I will, presumably, have a field of conflicting motives. For the ultimate decision to come through reflective endorsement as the activity of me, the agent, the decision must be made by a reflective procedure which aims at the unification of my field of motivation. To do this, I need to set up a principle which gives structure to what I do in this situation in a way that helps to give overall structure to the way I do things in general. I need to try (the aim) to come up with a deciding principle – a principle of which motive should win out in the circumstances – that fits as well as possible with other motives and decisions I have regarding other situations.

I might, for example, decide that I should study now and put the movie off or skip it, because I need to get a certain amount of work done. Going to the movie would satisfy a motivating urge that does not have a lot of clout in the overall structure of my motives and decisions. Satisfying some earlier decision to get a certain amount of work done each day, or to work toward some amount for the week or the month, could bring together many of the motivational pulls that help form a cohesive structure of decision-making from among my motivational field. This means that I am more unified by staying with my work than by going to the movie. Of course, there could be situations in which the movie were the more correct choice, but things would have to be different for that.

Importantly, it is simply aiming at unity that makes the reflective activity the activity of the agent, and the result, the action of that agent. This is the defining mark of the reflective procedure of motivational selection which plays the role of the agent and so

constitutes the activity of the agent. *Successfully* unifying oneself so far as possible is not required for the action to belong to the agent, that is what is needed for right action, or so my practical normativity argument would imply. The agent makes a decision and performs an action if it delivers a principled decision based on reflecting and aiming to unify the agent's motivational field. So, the agent does its job well, if it unifies well. Success as an agent is success at unification in decision-making. So, the reasoning goes, action is successful or right if it follows from success at unification in the decision that spawned it.

Therefore, we can conclude for now this sketch of the defining activity of the agent.

Unity Thesis: *For a principle to derive from an agent's reflective endorsement is for it to come from deliberation which aims at the unity of the agent.*

VII. Remaining Premises

A. Deontic Kantianism

I now argue for the remaining premises of the practical normativity argument. None of these arguments is extremely lengthy or involved, since they are already pretty well set up by the arguments that have already been made. I handle each of the premises in the order in which they appear in the argument.

In what follows I argue for the Deontic Kantianism Thesis, according to which right action follows from right deliberation and wrong action follows from wrong deliberation. In other words, it is the successful or unsuccessful performance of rational deliberation which determines whether the action as a whole is successful – that is,

whether the action is right or wrong. Since the standard for successful deliberation is set by the aim at unity, this means that right action is action which is done in accordance with unifying one's agency.

Since, as I have argued, it is reflective endorsement which makes the agent's contribution to action, which thereby makes it a true action by the agent, it is reflective endorsement, which is what the agent does in action. Rational deliberation is the process of using reflective endorsement to consider options and arrive at the agent's decision of what to do. So, deliberation can also be regarded as the activity of the agent. The fundamental idea is that the activity of the agent is that which performs the role of the agent. The role of the agent, in turn, is to mediate between desires and beliefs in order to come to a decision of what to do. This process is the activity of the agent.

Right action is action well done. Since the activity of the agent in an action is rational deliberation, it is rational deliberation which must be done well for the agent's action to be right. The part the agent does is the part the agent can do well or do badly. So, right action is action with its component deliberation done well. Conversely, wrong action, which is action done badly, is action with its component deliberation done badly. Therefore,

Deontic Kantianism: *Right action follows from right deliberation, wrong action follows from wrong deliberation.*

B. Attribution

The argument for the Attribution Thesis is easy to make, since it is aided by the very loose distinction I have made between an action by an agent proper and a mere act attributable to an agent. The Attribution Thesis defines mere acts attributable to agents as

behavior caused by the actions of agents proper through the agent's own attributable training or habituation to perform the act. Thus, acts which are attributable to agents are all acts caused by the reflective endorsement of the agent which issues a principle determining the act. The difference between an action proper and a mere act attributable to an agent is simply a difference in how direct the causal relation between the deliberation and the behavior is.

Talbot Brewer, in the context of arguing for a less stringent conception of a Kantian maxim (a point very much related to the Agent-Causation Thesis), uses an example he takes to be the action of an agent to show that actions are not always done, in my terms, on a principle (Brewer 2002, 544-545). The example is this: I might get a phone call in the middle of the night which I answer without really waking up all the way. In such a case, I certainly have not reflected and rationally deliberated on whether to answer the phone, but, according to Brewer, this is an action of mine.

I agree that this, at least to an extent, would be an act attributable to me. The extent to which it is attributable to me, however, is the extent to which I led myself to answering the phone through various forms of habituation which were themselves attributable to me. We can assume that I answer the phone with absolutely no reflection or deliberation at the time that I do it. However, that is not the only way for an act to be attributable to me, because I might train or habituate myself to act in such a way. The act might come from my character, which is attributable to me to the extent that it is my own construction, coming from deliberate decisions. For example, it might result from ways in which I have trained myself to respond to the phone ringing, including responding to calls at night, which stand a good chance of being important, or at least interesting.

However, there is a fuzzy line here which depends on direct versus indirect causal chains. We can compare the midnight call case to a clear case of action proper, like deciding on whether to go on vacation after days of careful deliberation. In both cases my behavior is caused by my rational deliberation – my reflective endorsement issuing in a principle of action. The difference lies in how immediately the act follows the decision. In the case of going on vacation, we may say that the act immediately follows the decision, in that I accept the idea and begin preparing to go, as soon as I have decided to. In the case of answering the phone, the act does not immediately follow the decision, since the decision was not deliberately aimed at this very act, but instead was a matter of casually, occasionally thinking about phones ringing, etc. (Rational deliberation does not have to be dramatic.) So, the distinction between these two senses of acts being attributable to agents, although it might be useful in some ways, is not fundamental and sits atop a commonality between in the two senses in what makes them acts attributable to agents. Therefore,

Attribution: *An act is attributable to a given agent if and only if it is an action that agent is the author of or it is behavior caused by habituation, training or preemption by (an) action(s) that agent is the author of.*

C. Transitivity

The Transitivity Thesis, according to which the rightness of any mere act attributable to an agent amounts to the rightness of the action proper which precipitated it, really follows from the same argument used for Deontic Kantianism. Since rightness flows from deliberation, and attributability flows from deliberation, rightness and attributability are bound together. Just as I argued for the Attribution Thesis, the

distinction between actions proper and mere acts attributable to agents is only superficial. In both cases, the act is caused by the deliberation. Similarly, in both cases, the rightness of the act follows the rightness of the deliberation. The deliberation which leads to the act determines the rightness of the act. Therefore,

Transitivity: *A mere act attributable to a given agent is right or wrong in accordance with the rightness or wrongness of the action(s) from which it derives its attributability.*

D. Assessability

The Assessability Thesis, according to which it is precisely acts attributable to agents that are normatively assessable in a strict sense, is a core thesis for constitutivism. While both the 'if' and the 'only if' components of this thesis are very important to the constitutivist line, at this point in the argument, it is really the 'only if' component that needs additional argument. That is, it is the restriction to acts of agents that still needs to be argued for, since setting up a standard of right and wrong for the acts of agents is the whole point of the practical normativity argument.

If some activity is not an act attributable to an agent, then it may be desirable or undesirable like any inanimate object. It may not, however, admit of rightness as something done well. It does not admit of being evaluated as an action. Just as we say that the behavior of viruses, worms and cows does not admit of normative evaluation, neither do any behaviors of humans which are not acts attributable to agents. The reason the behavior of cows does not admit of normative evaluation is because it is not the action of an agent. It does not result from rational deliberation, that is, reflective endorsement. It does not have the right structure to admit of normative evaluation. It is first-person deliberation that right action characterizes. The assessment of first-person

deliberation is what distinguishes the right from the good. Although we can think of anything at all as good or bad, in the sense of desirable or undesirable for some further purpose, it is only acts resulting from first-personal rational deliberation that can be done rightly. If something is not an act attributable to an agent, then it cannot truly be right or wrong, because it does not come from rational deliberation. Thus, third-person, theoretical assessments of actions are not applicable in a strict sense to activities or behaviors that are not acts attributable to agents.

If some activity is an act attributable to an agent, then it does admit of normative evaluation, because it then follows from rational deliberation, which can always be normatively assessed. Reflective endorsement, the cornerstone of rational deliberation, is the capacity to mediate one's disparate desires and motivational influences. It is the capacity to determine what one should do. This puts one in the realm of normative evaluation. The whole point of the constitutive aim arguments is to show that true rational deliberation can be right or wrong in an unqualifiedly normative sense, to use Stephen Darwall's phrase (Darwall 1992). If something is an act attributable to an agent, then it follows from the preceding premises that it can be right or wrong. Therefore, **Assessability**: *An act is normatively assessable if and only if it is an act attributable to some agent(s).*

VIII. Summary

I have now defended all of my major premises for the argument up to this point. Let us look again at the practical normativity argument.

1) **Agent-causation**: For an agent to be the author of an action is for the agent to act on a principle which derives from that agent's reflective endorsement.

2) **Unity Thesis**: For a principle to derive from an agent's reflective endorsement is for it to come from deliberation which aims at the unity of the agent. (To be replaced in Chapter 3)

3) Thus, if an agent authors an action, then that action derives from deliberation aimed at unity. (From (1) and (2))

4) **Deontic Kantianism**: Right action follows from right deliberation, wrong action follows from wrong deliberation.

5) **Attribution**: An act is attributable to a given agent if and only if it is an action that agent is the author of *or* it is behavior caused by habituation, training or preemption by (an) action(s) that agent is the author of.

6) **Transitivity**: A mere act attributable to a given agent is right or wrong in accordance with the rightness or wrongness of the action(s) from which it derives its attributability.

7) Therefore, if an act is attributable to a given agent, then it is right or wrong in accordance with its compliance with the aim at unity. (From (4), (5), (6) and (7))

8) **Assessability**: An act is normatively assessable if and only if it is an act attributable to some agent(s).

9) Therefore, if an act is normatively assessable, then it is right or wrong in accordance with its compliance with the aim at unity. (From (8) and (9))

I have, thus, argued for a theory of practical normativity and practical rationality. In doing so, I have argued for a specific theory of right action – that action is right or wrong in accordance with how well it (the deciding principle which is a component part of the action) unifies its agent. It is true that I have only offered a sketch of an explanation of agent-causation, which means that I am relying somewhat on the assumption that people do, in general, perform actions and acts attributable to them in the senses I have defined. I take this assumption to be relatively safe given that rational deliberation seems to be an important part of people's lives, a mark that distinguishes a special and peculiar thing that most people do from the rest of animal behavior (including various human behavior that does not make up action by an agent), and a way that people train themselves. Whether or not people do, in fact, rationally deliberate and perform actions in the way I specify is an empirical question.

In this chapter, however, I have not given my official account of rational deliberation, but only a sketch of it. It is the task of the next chapter to fill in my more detailed and formal account of rational deliberation. Doing so allows me to give a final version of the practical normativity argument. It also allows me to elaborate more accurately on how this argument and this account should be interpreted. After arguing for the Deliberation and Systematization Theses, I explain how systematization must work and what the picture of rational deliberation is that I am advocating. Doing this completes my account of practical normativity.

Chapter 3: Systematization and the Standards of Rational Deliberation

I. The goals of this chapter

In the last chapter I argued for a version of constitutivism, the view that practical normativity is grounded in certain constitutive aims of agency. I argued for a series of claims, which jointly formed my overall argument for the conclusion that insofar as an act is normatively assessable, it is right or wrong in accordance with its compliance with the aim at the unity of the agent. While I made this argument very clear and explicit through developing the supporting claims, I did not clearly draw out the implications of the conclusion, which remained just a sketch, due to a premise which was only a sketch or an outline of more detailed premises to be filled in. The premise, which brought the idea of the unity of the agent into the picture, was a sketch of the kind of claim that needs to go in its place. That place still needs to be filled in with something more precise.

Moreover, since I left the conclusion somewhat loose, I did not attempt to state what requirements were implied by it for an agent making a practical decision. The conclusion that an act by an agent is normatively required to comply with an aim at the unity of the agent, where the unity of the agent is a matter of unity in decision-making, is relatively vague and open-ended. Pursuing unity in decision-making through one's decisions does not have any one obvious course or method. It could involve repetition of similar acts, or suppression of weaker desires, or perhaps even valuing humanity as an end in itself, as perhaps Korsgaard might hold. The idea of unity as the guiding standard

of decision-making is a powerful one, but it needs to be further specified and filled in, so that the method of rational deliberation dictated by the aim of agency is clear.

In this chapter I continue to develop my theory of practical normativity and practical rationality by filling in the sketchy parts of the argument from Chapter 2, which brings me to my full, formal argument for my view. This, in turn, allows me to make more explicit what the requirements of practical decision-making are. It allows me to provide a fairly clear and potent set of rules for making rational decisions. Providing such a set of rules for making rational decisions is the ultimate goal of this and the previous chapter.

Before I begin making these arguments, I briefly compare my overall metanormative theory of practical rationality as systematization with a few similar views from the literature, like that of John Rawls, Sharon Street and Michael Smith, and I discuss the important differences. I intend this to help provide a sense of what my arguments are supposed to show and what the consequences of my general argument may be. In large part, I point out how my view does explanatory work that these other views do not, and how that helps form the view into the kind of metaethical theory that can withstand various plausible skeptical challenges.

I pursue the goals of this chapter in two stages. First, I replace the appeal to unity in my overall argument with the claim that, roughly, systematization of one's principles of action is the guiding standard for rational decision-making. This involves arguing at length for two new premises in the general practical normativity argument. Second, I explore the consequences of the new version of the argument and I use the new conclusion to derive a standard for how deliberation operates rationally. As part of the

investigation of the argument's consequences, I look at how our experience of what we usually consider rational deliberation, or a standard picture of what rational deliberation is, compares to what my argument implies and how this relationship can be plausibly understood. In other words, I compare my model of (correct) rational deliberation with what is often thought of as the model of (correct) rational deliberation.

II. The structure of rational deliberation

A. Transition from unity to systematization

In Chapter 2 I argued for the Unity Thesis and used it as one of the crucial premises in my explanation (or explanation sketch) of practical normativity and practical rationality.

> **Unity Thesis:** *For a principle to derive from an agent's reflective endorsement is for it to come from deliberation which aims at the unity of the agent.*

In this chapter I intend to develop that argument further so that it can be interpreted as a robust account of right and wrong action. In order to do this, I explain and replace the Unity Thesis with two other claims that have greater specificity and determinate content. These two claims take the idea of unity and turn it into a more specific phenomenon – systematization of commitments. By 'commitments' I just mean accepted principles of action (in the sense of principles of action developed in the previous chapter), where there can be different senses in which one accepts a principle of action, which I discuss later on. With the two new claims that replace the Unity Thesis the argument for my account of practical normativity can be stated in full, and from it I can derive the rules of practical reason which dictate the normativity of action. The first claim moves from what is necessary for rational deliberation to the need to draw from structured commitments, or

principles of action. The second claim moves from the need to draw from commitments to the requirement of aiming at systematization of commitments. These claims will be argued for at length.

> **Deliberation Thesis:** *For a principle to derive from an agent's reflective endorsement is for it to be drawn from that agent's set of commitments.*

> **Systematization Thesis:** *For a decision to be drawn from an agent's set of commitments is for it to come from deliberation which aims at systematization of that agent's commitments.*

In the first part of this chapter I offer direct arguments to substantiate these two claims. The more intuitive idea behind these claims is that making a unified rational decision requires formalization of one's various desires and dispositions so that they can come together to form a single perspective of the agent. This unity requires, first, that decisions are taken from formalized practical propositions, that is, principles of action; and, second, that decisions be formed from the principles by a formal unifying procedure, that is, systematization. The use of formalization allows the process to be rational in an explainable way, rather than by appealing to non-reducible rational intuitions or acts of will. In this Chapter, I present an explanation of the unity of the agent's perspective that simultaneously sets up a standard for right and wrong action.

While I offer the Deliberation and Sytematization Theses as an analysis of the Unity Thesis from Chapter 2, I do not argue for an equivalence between the Unity Thesis and the combined Deliberation and Systematization Theses. Instead, I simply argue that the two new theses are true. Furthermore, in order to ensure the strength of these two

new claims I do not rely on the Unity Thesis to support them. I argue for them on their own, independently of the arguments I used in Section VI of Chapter 2 for the Unity Thesis. However, I hope it will be clear that this new version does present an account of the Chapter 2 version and that unity is an initial way to understand what is going on in the systematization arguments.

With these two new claims in place, I am able to use the theses from Chapter 2 to provide my full, formal argument for my account of right and wrong action. This account, as the argument makes clear, is based on a theory of deliberative rationality embodied by some of the central premises. My formal argument is as follows.

Argument 7

1) **Agent-causation**: For an agent to be the author of an action is for the agent to act on a principle which derives from that agent's reflective endorsement. (From Chapter 2)
2) **Deliberation Thesis**: For a principle to derive from an agent's reflective endorsement is for it to be drawn from that agent's set of commitments. (To be argued below)
3) **Systematization Thesis**: For a decision to be drawn from an agent's set of commitments is for it to come from deliberation which aims at systematization of that agent's commitments. (To be argued below)
4) Thus, if an agent authors an action, then that action derives from deliberation aimed at systematization. (From (1), (2) and (3))
5) **Deontic Kantianism**: Right action follows from right deliberation, wrong action follows from wrong deliberation. (From Chapter 2)

6) **Attribution**: An act is attributable to a given agent if and only if it is an action that agent is the author of *or* it is a mere act caused by habituation, training or preemption by (an) action(s) that agent is the author of. (From Chapter 2)

7) **Transitivity**: A mere act attributable to a given agent is right or wrong in accordance with the rightness or wrongness of the action(s) from which it derives its attributability. (From Chapter 2)

8) Therefore, if an act is attributable to a given agent, then it is right or wrong in accordance with its compliance with the aim at systematization. (From (4), (5), (6) and (7))

9) **Assessability**: An act is normatively assessable if and only if it is an act attributable to some agent(s). (From Chapter 2)

10) Therefore, if an act is normatively assessable, then it is right or wrong in accordance with its compliance with the aim at systematization. (From (8) and (9))

The last two steps may seem superfluous, but they ensure the generality of the theory of normative standards for action. The Assessability claim was argued for in Chapter 2 to make it clear that there is no space between what acts are normatively assessable and what acts are attributable to the agent's whose bodies performed them. That is, the Assessability claim supports the whole constitutivist approach to explaining practical normativity (i.e., that normativity springs from constitutive aims of agency) by ensuring that there are no normative issues that cannot be handled through constitutivism.

The argument in its current form concludes with a claim about systematization and makes it the guiding law for all action. This is a more explicit and specific

conclusion than the Chapter 2 conclusion that actions are right in accordance with the aim at unity. From the aim at unity, it is still very much unclear how any particular rules of practical thought and action could be derived. With systematization, however, a relatively clear and profound set of rules can be established. Deriving these rules from my arguments for systematization is what I will do in the latter part of this chapter. In doing so, it will also become clearer how unity is still a relevant concept for understanding the rules of systematization.

B. Comparison to similar views

This theory of rational deliberation I am arguing for and expounding is similar to a few theories concerning practical reason and morality which are already in the literature. Perhaps the most obvious similarity is with theories that draw on reflective equilibrium, as introduced by John Rawls (Rawls 1951, 177-197) and developed by Norman Daniels (Daniels 1979, 256-282). To an extent, my theory can be thought of in terms of deliberation as seeking reflective equilibrium with practical principles. In line with this, one could apply some of the methods used in the discussion of reflective equilibrium to an understanding of how the systematization of commitments works.

There are significant differences, however. Reflective equilibrium is a procedure for deciding on beliefs. According to my theory, what one is supposed to systematize and what is supposed to be derived from doing so, are practical principles. Further, reflective equilibrium involves balancing and organizing all of one’s beliefs (or at least all of one’s relevant beliefs), which could involve a great deal more than evaluative beliefs. Since only evaluative beliefs are really comparable to practical principles, this is a large difference. What one is supposed to systematize, on my view, does not include

theoretical beliefs, which are not supposed to be adjusted to accord with practical principles (at least not unless there is some special relationship in a particular case).

Not only are there various technical differences between what is involved in using reflective equilibrium and what is involved in aiming at systematization, but there is a more fundamental difference in what the theories are about. Sharon Street provides an excellent discussion which makes it very clear how the kinds of theories put forward by Rawls and others are decidedly distinct from what I am arguing for (Street 2008). Street distinguishes between what she calls "restricted constructivism" and "metaethical constructivism." She defines:

> *Restricted constructivism in ethics* specifies some particular, restricted set of judgments about reasons, and says that the correctness of a judgment about reasons falling within that set is constituted by the judgment's withstanding a certain (specified) procedure of scrutiny from the standpoint of some (specified) set of further judgments about reasons. (Street 2008, 209-210)

Which is distinguished from:

> *metaethical constructivism*, [the theory according to which] the fact that *X* is a reason to *Y* for agent *A* is constituted by the fact that the judgment that *X* is a reason to *Y* (for *A*) withstands scrutiny from the standpoint of *A*'s other judgments about reasons. (Street 2008, 223)

As Street argues, Rawls theory is a version of restricted constructivism. It does not really attempt to cover the ground that my metaethical constructivist theory does. There might be something to be said about whether Rawls still needs a theory behind his

restricted constructivism, but even if Rawls does not need a more general theory of reasons, it remains true that he does not provide such a theory.

While Street does an excellent job of setting out metaethical constructivism as a clear metaethical view, she also offers a sort of minimalist version of metaethical constructivism as her own view. According to her, there is nothing more to the correctness of reasons or values beyond their agreement with each other for a particular agent. She also cites a kind of standard of authority between normative judgments in that those which are "closer to the core of a person's interlocking web of normative judgments" are more truly the agent's own, and so are claimed to have more authority (Street 2008, 235, n. 45). This is very much the kind of view that I am arguing for and expounding. Other than her not providing any arguments or explanations for her view (the point of the paper is to clearly define the view and distinguish it from related views), I really only differ from what she says in terms of specificity.

The theory of normative reasons from Michael Smith also has significant similarities with my theory (Smith 1994; Smith 1995, 109-131; Smith 1997, 84-119), but it too is different in its technical aspects of how it works and in terms of what it is about. Smith proposes that having a normative reason is in part constituted by the rational demand to systematically justify one's desires. I argue later in the next subsection, II.C, that Smith's theory does not work as an explanation of how deliberation operates as a rational process to form a principle of action. That criticism depends on arguments I made in Chapter 2 and only shows that what Smith says does not fit into my theory. Now I would like to offer a more general comparison and criticism of his view as a full blown alternative, although I still only examine what he says that is similar to my view.

First, I would like to point out that my scheme is much simpler than Smith's in terms of how it is supposed to work. On my view, there are possible commitments and we compare different ways to systematize them. All the relations to examine are structural relations between the principles. We look at which set of potential commitments has the greatest systematicity, which will naturally bring us to include some potential commitments and exclude others. This only requires looking at how the commitments relate to each other – whether they come from our motivations is already given in virtue of them being potential commitments.

According to Smith's view, lots of things must be taken into consideration in working out the systematic justification of desires. Desires are supposed to be justified in a way that involves checking the theoretical beliefs they are based on or relate to *and* connecting the related beliefs and the desires through systematic relations. This involves adjusting desires so that the desires the agent has fit in better with the agent's other desires (Smith 1994, 159-61).

It is important to note that Smith has no explanation of why systematically justifying desires is rational. He just claims that it is. This might be an intuitively appealing claim, but it seems inappropriate for him to leave this unexplained. First of all, this cheapens his general explanation of how truly normative reasons can be squared with the empiricist claims of judgment internalism[10] and (more importantly) the Humean theory of motivation, according to which motivation consists of desires and means-end beliefs. Second, this weakens the plausibility of his claim that we have good reason to believe that "there will be a convergence in our desires under conditions of full rationality" (Smith 1994, 187). The plausibility of the claim is weakened, because

[10] See Chapter 1, Section VI.B.

without a general explanation for why systematically justifying desires is rational we have no specific standard for the justification of desires that would indicate that everyone's desires will converge in accordance with the standard. Of course, the "empirical evidence" Smith appeals to for the claim that conditions of full rationality lead to a general convergence of desires is not very inspiring (tractability of moral disputes, our tendency to increase agreement with moral discussion, etc.). So, what we really need is some reason to think that this convergence is determined by the standards for systematic justification themselves. Independent of any explanation of the rationality of systematic justification, there is no such standard to be found. So, Smith's claim is left looking rather weak.

The lack of a general account of practical rationality which explains the need to systematically justify desires also weakens Smith's position with respect to the following problem. Desires, despite Smith's claims to the contrary, cannot be systematized themselves, because desires are not the kind of formal, structured thing that can be systematized. Consider for example his claim in regard to getting rid of desires that do not fit in with one's other desires:

> Perhaps because we can see no way of integrating these desires into the set as a whole they will come to seem *ad hoc* and so unjustifiable to us. Our belief that such desires are *ad hoc* may then cause us to lose them. And, if so, then it will seem sensible to describe this as a loss that is itself mandated by reason; as again straightforwardly analogous to the loss of an unjustifiable, because *ad hoc*, belief. (Smith 1994, 160)

But how can a desire be *ad hoc*? A desire is just part of who a person is. That is why there has been the whole controversy in the history of moral theory over whether moral imperatives depend on desires. Desires, by themselves, are not the kinds of things that can exist for a good reason or be *ad hoc*. It may be true that a desire seeming to be *ad hoc* to a person will cause the person to lose the desire, but, aside from the plausibility of this claim, this can hardly be called a loss "mandated by reason." If I want x, and wanting x comes as a surprise, but I continue to want x nonetheless, there is nothing, rationally, wrong with this. Having an *ad hoc* desire is not any kind of rational violation or failure. I can desire pea soup, even if this does not fit into any systematic scheme of beliefs and desires, all I want without violating a dictate of reason or coming up short rationally.

The problem is twofold: (1) desires are not the kind of formal things that admit of systematization, and (2) Smith has no general explanation of practical rationality, and no explanation of why desires should be systematically justified. Smith could get out of this with a general explanation of practical rationality. With this, he could talk about rational changes in attitude, which cause changes in motivation. This could be called a change of desire through rational process; however, as things stand, this is not available to him.

My theory, on the other hand, includes such a scheme. Changes in principle acceptance in accordance with overall systematization are rational, because this is exactly what the work of practical reason is, as I have explained. Changes in principle acceptance of course may cause changes in motivation, since the principles are formed from one's dispositions to be motivated. Such changes in motivation would be

"mandated by reason," because they follow from the conformity to a rational requirement.

C. An analysis of deliberation

In what follows I sketch out a theory of rational deliberation and decision-making. I argue for this theory by using the conclusion from Chapter 2 that rational decisions, the decisions *of agents*, must be made on principles of the form "I do X, on condition Y." These principles give structure to the motivational attitudes present in the individual and make them something that belongs to the agent. I continue with this line of reasoning by arguing that these principles are not just the result of deliberation, but they are also what deliberation makes use of in order to come to a result. In this subsection I argue for the Deliberation Thesis and in the following subsection I argue for the Systematization Thesis.

To begin the argument for the Deliberation Thesis, let us consider the role of motivation in rational deliberation. Everyone has a set of dispositions to be motivated in various ways. This set of dispositions is determined by a number of psychological factors including one's desires, one's beliefs, one's intellectual abilities, one's motor skills and so on. As I said in Chapter 1, I am arguing for a theory that takes motivations to be a constituent part of what one has normative reason to do, and what one is obligated to do – I am arguing for a version of internalism. In what follows, I argue for the Deliberation Thesis and the Systematization Thesis in ways that begin from an understanding of how one's motivations factor into one's deliberation and practical reason. I start by talking about one's motivational dispositions and how these relate to practical rationality via their role in rational deliberation. On my view, motivational

dispositions are the foundation of rational deliberation and so have a direct and fundamental relationship with practical reason and obligation. In a way, my view is even more internalist than most internalists, because I do not merely claim that one is motivated when one is rational and then tell a different story about the standards for being rational. I claim that motivation is the basis for what it is to be rational.

To get my account of deliberation started, I first need to set up the place of motivational dispositions. While motivational dispositions (or attitudes like desires) play a fundamental role in rational deliberation, they cannot do it alone. Recall that, as I argued in Chapter 2, mere motivational dispositions cannot explain an agent's action, in the sense of an action performed by an agent proper. The reason for this is that motivational dispositions cannot account for the agent's ability to intervene between them and contribute to the process that leads to bodily movements, in Velleman's terms (Velleman 1992, 461-481), or they cannot provide a perspective that encompasses the whole field of motivations rather than just the perspective internal to each motivation, in Darwall's terms (Darwall 1983, 103). The general idea is that the interaction of motivational dispositions as such cannot add up to the rational process of an agent. The decision-making process of an agent must have the form of normative reasons and justification. Since reflective endorsement can explain action, motivational dispositions alone cannot explain reflective endorsement, because otherwise they could thereby explain the action of an agent.

This leads to the issue of what can fix the problem. The short answer is adopting structure for one's motivations. One needs to adopt or impose some form of structure for one's dispositions which can get them in the right shape to add up to a rational process.

Doing so, adopting structure, is taking them to be something more than they otherwise are. To use the language of theoretical inference, the procedure of adopting structure is "ampliative" just as induction and inference to the best explanation are ampliative forms of inference. Induction concludes more than what is strictly contained in the evidence itself; whereas, deduction does not conclude more than what is contained in the evidence and so is not ampliative. Since adopting structure for one's motivations is ampliative, one's complete set of motivations, or motivational dispositions, does not strictly determine the structure that belongs with it or on it, even when it is decided what kind of structure is in question. I say more about the openness of amplifying one's motivations with a given kind of structure below.

Deciding what *kind* of structure is to be adopted is an important issue in its own right. For the purpose of how deliberation must operate, it is the forming of practical principles, or *commitments*, which must be imposed on one's motivations by the agent. In Chapter 2, I argued for the Agent-Causation Thesis that being the author of an action is the agent acting on a principle which derives from the agent's reflective endorsement. This notion of a principle is a practical proposition of the form, "I do X on condition Y." An important feature of practical principles of this kind is that they have universal form. By reflectively endorsing an act on a given principle the agent adopts a rule of what the agent does anytime the condition arises. Of course, agents will need to change and adjust their rules as they go along in life, but the rules themselves have universal form, which means that the agent goes against the agent's own will if that agent acts contrary to an adopted principle without changing it. In this sense, principles are *commitments*. In order to emphasize this feature and to help make some of what I say more intuitive, I use

the term ‘commitment’ rather than ‘principle’ in most of this chapter’s discussion, although I always use these terms as synonyms unless I explicitly state otherwise.

I have already argued that one needs to adopt some form of structure for one’s motivations in order to make deliberation with these motivations amount to a rational process which embodies reflective endorsement and the activity of the agent. Unsurprisingly, I now argue that the kind of structure one must adopt is that of commitments, although at this point I only argue for a weak form of this. Later, I bolster this result with additional considerations.

The point of adopting structure for one’s motivations is to make them add up to a rational process which embodies reflective endorsement and thereby the agent’s contribution to a decision. Reflective endorsement, to do its work, must result in a principle, as I argued in Chapter 2. So, the specific thing the structured motivations in deliberation must add up to is a principle of what the agent does on a given condition. In order for this to work, whatever structure is imposed on one’s motivations, it must be done so that the result has bearing on what the agent does on various conditions.

I have already pointed out that imposing structure on one’s motivations is ampliative and so is not strictly determined by one’s set of motivational dispositions. On the other hand, I have argued that the agent must impose structure that has bearing on what the agent does on various conditions, which suggests a use of commitments. In order to meet the needs set by these two claims, we need to consider what I call “potential commitments.” One has a *potential commitment* that one does X on condition Y just in case one’s motivational dispositions could be consistently formed (in the logical sense) into such a principle of action that one does X on condition Y, where the consistency is

relative to other standing commitments and standing beliefs about what "principle-conditions" (the conditions cited by the principles) are compatible. In other words, one's set of potential commitments includes all the possible formations of commitments from one's complete set of motivational dispositions. This still allows potential commitments to be inconsistent with each other, since one potential commitment does not depend on what other potential commitments one has. I could have motivations such that I have a potential commitment to spend all my time playing games when I'm not eating or sleeping, and another potential commitment to do philosophy at least six hours a day. These potential commitments are not compatible with each other, but they are both potential commitments for me.

So far, there are only a few things that restrict the field of potential commitments for an individual. One cannot have a commitment to do something where one has no motivation to do that thing. I cannot have a commitment, that is, I cannot form the practical principle, "I ride my bike, when it's raining," if I have no disposition to be motivated to ride my bike when it's raining. Of course, dispositions are complicated things. It is unlikely that I have no disposition *at all* to be motivated to ride my bike when it's raining, since there may be some situation in which something I deem very important depends on my doing so. However, I do have some dispositions and not others which makes for certain restrictions on potential commitments. My dispositions depend on many details rather than just some simple condition, but this does not leave the field of potential commitments entirely open. I may have a disposition to ride my bike in fair weather if my muscles feel good, but a disposition to *not* ride my bike in fair weather if

my muscles do not feel good. Nevertheless, I still have dispositions to bike and not bike on certain conditions, which are compatible with some practical principles and not others.

So, one is restricted to that with which one's motivations are compatible, but since motivations do not come pre-structured, so to speak, this only restricts potential commitments to those one has some motivation to support. Later, I elaborate on the fact that some potential commitments are incompatible with each other which means that there will be various sets of potential commitments which must be treated separately from each other. This, as well as other factors, produces more restrictions on potential commitments. Regarding what I have argued for thus far, however, the only additional consideration relevant for one's potential commitments is how one's beliefs affect one's recognition of various conditions. For example, I may from time to time see someone unknowingly drop his or her wallet. I may in fact be disposed to be motivated to give back dropped wallets, but this may only be triggered when I believe that I will get some kind of reward for doing so. If I am not altruistically inclined enough, a belief in reward may be required to trigger willingness to return the wallet. The dependence on my belief in the reward is held as a constituent part of the disposition. This restricts potential commitments for me concerning returning dropped wallets. Although this point can be generalized to many kinds of dispositions, it is really just a point about how beliefs play a role in motivational dispositions.

Given what has been said so far, one's field of potential commitments is more or less wide open. Potential commitments do, however, display the structure of commitments. They are at least proposals for principles to be adopted and they are ways of representing one's dispositions as possible commitments. So, they are a step in

imposing structure on one's motivational dispositions, and this step has the features of being represented as practical principles. This is important for the reason that was stated above: the structure one imposes on one's dispositions must have bearing on what one does on various conditions. Of course, the most direct way to do this is by considering motivations as potential commitments. Considering motivations in this way is considering them as complexes which determine what the agent does on given conditions. That is exactly what is required by the need for bearing on what one does on various conditions.

The need for ampliative structure to be imposed by the agent on the agent's motivations does not directly indicate what kind of structure must be imposed. We may want to consider what other structures might be imposed to do the same work, and whether such structures might not be superior in some way or at least optional. For example, Stephen Darwall argues that we must make our preferences "criticizable by reasons" in order to make our agency coherent (Darwall 1983). Michael Smith argues that we must make our desires "systematically justified" (an idea very much related to what I say below) for us to be fully rational (Smith 1994, 159). These or other structures might be ways of making deliberation with motivations rational and a function of what the agent does – they could be what is missing from an explanation of the agent's practical decisions.

These apparent alternatives, however, are not viable alternatives at all. First of all, both of the example alternatives appeal to reasons or justification: normative phenomena we are trying to explain. More important though, these proposals do not get past the problem that what the adopted structure must do is create the needed outcome – a

practical principle – as a form of reflective endorsement. That is to say, the work of the structure is to provide a basis for what the agent does on various conditions; the work is to provide a basis for principles of action. Any structure that does this will, one way or another, form possible ways of organizing one's motivations into what the agent does on various conditions. This is the same as organizing one's motivations as a field of potential commitments. Hence, any structure one imposes on one's motivations that does the appropriate work will directly or indirectly amount to a field of potential commitments.

Having a field of potential commitments puts one in a certain position. It gives one a store of commitments, potential and actual, which can be used to deliberate and come to a reflectively endorsed decision. For the purpose of deliberation, one has at least one's field of potential commitments. One also has the actual commitments which one has explicitly accepted or which are implied by what one has explicitly accepted as a part of the decisions one has made. This store of commitments can be viewed in terms of how the commitments within it can fit together. Some commitments and groups of commitments will be compatible with other commitments, and some will not. There is much complexity in the store of commitments which will play an important role in the systematization model of rational deliberation.

So, every agent has a store of commitments, some explicitly accepted, some implied by what is explicitly accepted and some that are merely potential. The agent needs this store of commitments, because the agent needs amplified structure for the agent's considerations to amount to a rational process and the agent needs to adopt structure so that it has bearing on what the agent takes as a principle of action (the agent's

practical decision). The simplest and most direct way to do this is to consider our motivations as potential practical principles. Moreover, as I argued above, one way or another, the agent *must* adopt structure that has the form or the implication of what the agent does on various conditions. So, even apparent alternatives like "making preferences criticizable by reasons" (Darwall) or "systematically justifying desires" (Smith) will amount to the same thing in this context.

Hence, the deliberating agent needs to consider the agent's field of motivations *as* potential commitments. Recall that, as I argued in Chapter 2, the activity of the agent is to be equated with reflective endorsement – the procedure which plays the role of the agent by mediating between and selecting among one's motives through a reflective process. What form this procedure takes can be inferred from what is necessary for it to mediate between and select from among one's motives so as to constitute the activity of the agent. So far, I have argued that this procedure must involve imposing structure on one's motives so that they can be considered as potential commitments. Doing this in the context of rational deliberation – that is, reflective selection of motives through the appropriate procedure – is drawing from one's commitments. The reflective endorsement procedure must consider one's motives as a field of potential commitments and use this field to make the reflective selection. The procedure must be guided by the structural topology of the field of potential commitments. That is the whole point of imposing the structure – to consider one's motivations in such a way so as to provide a formal procedure for determining a principle of action. This is to draw from one's field of commitments. Therefore,

Deliberation Thesis: *For a principle to derive from an agent's reflective endorsement is for it to be drawn from that agent's set of commitments.*

It is true that I have argued for the Deliberation Thesis in way that makes it seem weaker than what one might initially expect. I have argued that one must draw from one's *potential* commitments rather than one's accepted commitments. I freely admit that this is a substantial difference. However, this interpretation of the claim is exactly the one I use in my argument for the Systematization Thesis, and this interpretation is the one I use when I discuss the practical import of the entire argument. As it turns out, drawing from one's potential commitments instead of just one's accepted commitments does not weaken my ultimate conclusion. In fact, it gets around what would be a serious problem: how to accommodate one changing one's mind and realigning one's commitments. If one had to draw from accepted commitments in order to deliberate rationally, one would not be able to rationally choose a new way of organizing one's motivations. It might, for example, be impossible for someone who had been raised with a conservative, religious outlook to rationally give up the practical side of this outlook and become liberal and non-religious (or vice versa). One's accepted commitments would force one to keep the same framework in order to deliberate rationally. In order to get out of this trap, one would either have to commit what would be large mistakes in deliberation, or undergo some kind of extra-rational, extra-agential transformation. Such an interpretation cannot be right. With the potential commitment interpretation, however, this problem can be easily addressed. Every decision is made with one's potential commitments, so it is always rationally open to reorganize and restructure one's

commitments along new lines. I discuss later on some pragmatic limitations on this open-endedness. An additional point in favor of this interpretation is that this interpretation, along with the corresponding interpretation of the Systematization Thesis, yields an account of practical reason that can be used to reach the goals of this project by providing a robust and well-working model of right and wrong action.

D. Systematization argument

For the complete argument for an account of practical normativity (the overall argument I am making in the previous and the current chapters) to reach a conclusion that has any hope of regularly providing determinate answers to practical questions we need an analysis of what it is to draw from one's set of commitments. Without such an analysis, the argument would conclude that right and wrong are matters of whether or not one draws one's decision from one's set of commitments, or perhaps whether one *successfully* draws from one's set of commitments. In either case, it seems that the conclusion is not in the proper condition to be practically *or* theoretically helpful in determining what the right thing is for a particular agent to do in a particular situation. If you are considering whether to steal or pay for a shirt you like that is hanging on a rack outside of a store, it would not be very helpful in determining what you should do, practically or theoretically, to know that your decision should be drawn from your set of commitments. This is not enough to nail down what right or wrong action is. What is needed to improve this situation is an analysis of what is involved in drawing from one's set of commitments. This analysis is given by the Systematization Thesis.

My argument for the Systematization Thesis will come in two stages. First, I argue for a two part claim: that *drawing from a body of material requires an aim and a relation of affinity between the thing drawn and the body drawn from.* For example,

drawing from the *Critique of Pure Reason* to create a theory of the nature of mathematical knowledge requires aiming at some relation of affinity between things said in the *Critique of Pure Reason* and what the theory is, like a relation of similarity of content or a relation of common method of argument. Drawing one's political opinion from a set of newspaper articles requires aiming at a relation of affinity between the opinion reached and some or all of the contents of the articles. The relation of affinity might be a kind of similarity in how the opinion and the contents of the articles agree with some background principles about what facts should or should not hold of a candidate, party or policy, which is being judged.

Second, I argue that *the particular relation of affinity required for the case of drawing from a set of commitments to produce a principle of action as a deliberative decision is systematization*. Composing these two claims, of course, yields the Systematization Thesis. So, the argument for this thesis depends on a general conception of what it takes to *draw from* a body of material, and what it takes in particular to draw from a set of commitments to produce a decision in a way which amounts to reflective endorsement, or an agent's act of will.

One may wonder what the significance is of pushing the explanation further and further down, with more and more complexity resulting from doing so. Why stop with systematization? Why not analyze and explain that in order to get a new level of analysis and a new result? The answer to these questions is that if we could keep going, that would only help. Stopping where we do is just a matter of how much of the structure of rational deliberation can be unearthed. If we could analyze further and reveal more of the structure, that would give us a more precise account of what the standards of deliberation

are. In Chapter 2 I argued for the Unity Thesis which gave us a certain level of the structure of reflective endorsement and acts of will. I do not think it is sufficient to stop there, because I do not think that the Unity Thesis has enough structure to give us a well-informed view of the nature of practical reason. If all we had to go on were that agents must aim at unity in their principles, I do not think we could discern a whole lot about what kinds of decisions would be good ones and what kinds bad. Even with the added analysis of systematization, there are still limits to the extent to which practical reason and reasons for action will be determinately specified. Nevertheless, this analysis goes some distance in giving a clarificatory (as well as explanatory) account of practical reason.

To return to the task at hand, I first need to argue that drawing from a body of material requires an aim and a relation of affinity. The first step in making the "draw from" argument is delineating the context in which we are discussing "to draw from." I need to be able to say that *in this context* to draw from requires an aim and a relation of affinity. This allows me to state a general rule under which my case falls.

If a conceptual object, or object with conceptual content, P is drawn from a set of conceptual objects, or objects with conceptual content, S (a "body of material"), let us call S the *drawing basis* for P and the object drawn the *drawn object*. For any kind of conceptual object, like a proposition, to be drawn from some other conceptual objects is for it to result from an aim at a relation of affinity with the drawing basis. There must be an aim, because this is what makes the drawing basis productive of the drawn object. Without an aim, the drawn object does not arise through being drawn from anything. The aim also must be at a relation of affinity. If there is no aim at a relation of affinity, then

the drawn object is not drawn *from* the drawing basis, but rather, if it is drawn from anything at all, it is drawn from something else. This can be seen by looking at a series of examples.

Say that the fictional character Aeneas is drawn from the character Odysseus. If this is true, then Aeneas must have been produced with an aim at some relation of affinity with Odysseus.[11] If there was no aim, but there simply exists some relation of affinity, then it is not true that Aeneas was drawn from Odysseus. Instead, Aeneas merely bears the relation, say, resemblance or common literary role. There could still be a historical relation between the epic characters, but the relation is not that of the one having been drawn from the other.

One might object that the drawing from relation is causal, and so does not require an aim: there is a causal relation (of a certain kind) between the characters. For instance, one might specify a definition of 'draw from' that relied on causation under "normal conditions." So, Aeneas is drawn from Odysseus just in case there is a causal chain from the creator's experience of Odysseus to his creation of Aeneas, under suitable conditions. This suggestion is very sloppy, but the point is that the draw-from relation may appear to be based on merely causal relations rather than intentional, aiming relations.

However, while this could be a relation of *influence*, it could not be a relation of *drawing from*. The draw-from relation cannot be "external," but must be intentional in the minimal sense I am suggesting. If the relation is causal without the aim, then the character was not drawn from, but influenced by. This distinction is very important. A relation of influence will not add up to an intentional relation unless it is of the right

[11] Although Odysseus and Aeneas may have been real, historical people, I am considering them as fictional characters. Whether the characters were based on real people and real events is its own issue. It does not change the fact that the fictional characters were created by one or more authors or storytellers.

particular kind. We may well classify the draw-from relation as a relation of influence, but obviously the converse is not true. Influence, in general, does not essentially involve intentionality, and intentionality is required for the draw-from relation.

In any event, it would seem that in our case one's decision as a rational agent being caused by one's commitments without an aim is impossible, since we otherwise return to mere motivational dispositions yielding the decision, which I have already argued is incompatible with rational deliberation and the decision of an agent. If there simply is a causal relation with no intentional structure to it, which transmits influence from one's motivations to one's decision, that is no different from simple unstructured motivational dispositions producing a response. This might be an integral aspect to deliberation and decision-making, but it cannot be the whole story, because there must be some kind of structure imposed on one's motivations, and, specifically, one must draw from one's set of commitments. Reducing the draw-from relation to a simple causal relation just brings us back to where we started – unstructured dispositions which cannot amount to a rational process.

One might also object that using examples of a literary author skews the issue and forces us to attribute an aim – a soft form of intentionality. However, a literary author is not required in the example to make the point. Take the deliberation of a jury. Say that it draws its verdict from the evidence presented in court. This requires an aim at a relation of affinity between the verdict and the evidence presented in court. This aim can be established by rules of jury deliberation and orders of the judge. In other words, the aim comes from background intentionality, and not necessarily the intentionality of the deliberating jury. Furthermore, although the relation of affinity in this case is very

complicated, there clearly must be one. What that relation is might, for example, be construed as synthesis or assimilation, with some other things built in, but the relation must be there and the aim at the relation must be there in order for the jury to be drawing its verdict from the court's evidence.

Take another legal example. The Supreme Court rules on a case. The ruling of the court is not equivalent to the ruling of any individual, even though some individual might write the majority opinion. The decision is drawn from past rulings because the process aims at a relation of affinity with those past rulings. Of course, this is in part similar to the literary author example, because each judge creates a decision with an aim at a relation of affinity (resemblance, agreement, consistency, continuation, extension or some such). However, that is not the court's ruling. The court's ruling is reached by voting (which has a process of its own) and a majority opinion is written. The result is the court's ruling, which is created with an aim at a relation of affinity with past rulings via a process of individual aiming, voting and consensus.

If this is not the case, then the court does not draw from past verdicts, even if there is some causal influence on the court's verdict from some past cases via persuasive influence, framing of the issue, or other historical causal relations. If some or all of the judges are influenced (for instance) by past rulings, but aiming at a relation of affinity with past rulings is not built into the process of reaching a verdict, perhaps because doing so is not part of the legal method in use, then the influence from the sources in question is no more relevant than the influence of how well the judges are sleeping or how their family life is going. Once again, an aim is necessary to draw from some body of material. A process can aim at a result, particularly a relation of affinity, and thereby

draw from some material just as well as an individual person can, at least it can if there is some basic background intentionality in the process. In fact, at least in part, the aims of a process make it the process that it is. The important thing is that the court's proceedings can and *must* have an aim in order to draw from a body of material to arrive at its decision.

To further press the point about background intentionality, let us also look at the example of a committee trying to produce a document. A committee's decision or a document it produces may draw on various other documents. This is not reducible to a single member drawing on the documents or the committee members "on average" drawing on them. The committee drawing from the documents is also not equivalent to a rule behind the actions of the members which explains their behaviors. The committee draws on the documents because of the way the committee is set up, the task that has been set for it (either set by itself, or by an outside body – perhaps the committee was created for the purpose in question), and the particular activities of the members, which have their own purposes or patterns. If a committee is supposed to create a pamphlet outlining the essence of some event, which is supposed to draw from past pamphlets for the event, pamphlets for similar events, etc., one can imagine various ways this might happen. Members work on individual parts, make suggestions, give criticism, debate what should go where and so on. Some members may want to break from the old documents while some members may want to copy from them.

Insofar as the committee does draw from old pamphlets to create the new one, it must have an aim at a relation of affinity between what the committee creates and the old pamphlets. If the committee does not *aim* at similarities, common themes, common

styles, common lineages, synthesis of the old pamphlets as a collection, or some such then whatever affinities the new pamphlet has with old ones, the committee did not *draw from* the old pamphlets in creating the new one.

Moreover, if the committee does not aim at *affinity* between old and new, then it obviously does not draw from the old to create the new. There must be a relation of affinity between the drawn object and the drawing basis, which is aimed at, because that is what makes it the case that the drawing is *from* the drawing basis. If the committee looks at the old pamphlets as what not to do, then the committee may be using the old pamphlets, it may be learning from them, or it may even be drawing from what it learns from the old pamphlets, but it does not draw from the actual content of the old pamphlets themselves. Both an *aim* and an aim at a relation of *affinity* are required.

I am not sure whether an aim can result purely from a procedure with the background intentionality explained by the procedure and is reducible in this sense. This is a complicated issue which is taken up in other areas of philosophy. My purpose is not to explain intentionality all together, but just rational agency. Rational agency does presuppose a kind of limited intentionality, but that is not circular. Restricted forms of intentionality can explain the grandiose form of intentionality that is rational agency.

To continue with my analysis of rational agency, I need to argue for the second step toward the Systematization Thesis. I have already made a considerable fuss about the importance of structure to the process of deliberation. The first way in which this enters the picture is by considering one's motivations as a collection of potential commitments – in a sense, as a collection of policies about what to do in various situations. I then went on to argue that this structure must be maintained and utilized by

drawing from it, in the sense that one must aim at a relation of affinity between what is drawn and what is drawn from. I now want to argue that there is only one way to construe the relation of affinity in question, such that this structure is preserved and exploited. The relation must be systematization.

There are, of course, many possible relations of affinity. There are various kinds of content and surface similarity. There are various kinds of common purpose or common source. There are also various kinds of structural affinity. It is this last variety which must take precedence over the others for the particular case at hand: how deliberation draws from one's potential commitments to arrive at a principle of action. The structure of one's potential commitments is vital to the process which embodies an agent's deliberation being a rational process. If one's deliberation aims at a relation of affinity with one's set of commitments which is non-structural, like similarity of the act and the condition, then the structure of the commitments cannot really be doing the work in arriving at the new principle. Instead, the work is being done in an analogous way to how the potential commitments were formed in the first place. There is a rather arbitrary formation of a principle in connection with particular dispositions to be motivated. This cannot be the rational process of the agent, because the agent's base of decision-making – the agent's set of motivations – has not been brought together. The whole point of considering motivations as potential commitments and drawing from them is to do precisely that, to bring the agent's set of motivations together to come to a decision. If the relation of affinity aimed at in the deliberative process is non-structural, then the capacity of the adopted structure to bring the agent's motivations together is lost. Bringing the agent's motivations together is just what has to happen in rational

deliberation so that the decision can be the agent's own in the required sense. Hence, for rational deliberation the relation of affinity cannot be non-structural – it must be structural.

If a structural relation of affinity is required by the rational deliberative process, what must that relation be? The answer is simple. It is the relation which encompasses all structural relations of affinity between propositions: systematization. Propositional structural relations include compatibility, implication, and system unity. Since it is simply a structural relation of affinity that is needed, the relation must be all-inclusive. That might be bad news if the relation had to encompass all similarity relations, because it seems like such a generalized similarity relation would be rather indeterminate. This is not the case for the all-inclusive *structural* relation of affinity. While systematization may have limits on determinacy, it is largely sufficient to determine the nature of the arrived-at principle of action, the decision, from the basis of one's set of commitments, but I will say much more about this in the next section.

While I will explore the full implications of this in Section III, it is already clear that in order to meet the stated requirements for the process that leads to a principle of action to amount to an agent's rational decision the following two things must be the case. First, the process must aim at a relation of affinity between the drawing basis, the agent's set of commitments, and the drawn object, the principle of action which results from the process. Second, the relation of affinity aimed at must be systematization, the all-inclusive structural relation for propositions. Putting this together,

Systematization: *For a decision to be drawn from an agent's set of commitments is for it to come from deliberation which aims at systematization of that agent's commitments.*

III. The consequences of the argument for systematization

A. How deliberation as systematization works

I have now argued that *if an act is normatively assessable, then it is right or wrong in accordance with its compliance with the aim at systematization*. This is the summary claim of my theory of practical normativity. In order to make clearer what my theory involves and to set up later arguments for its applicability to the foundations of morality, I next explore the implications of this summary claim. Specifically, I discuss how rational deliberation works and what its standards for success are. In the following subsection, III.B, I discuss the relationship between the formal standards for rational deliberation and our experience as deliberating agents who make decisions involving normativity.

The aim at systematization, which sets the standard for correct deliberation and right action, is an aim at putting all of one's commitments – that is, principles of action – in systematic unity. As I discussed in Section II, one's commitments are rather fluid in the sense that, strictly speaking, one always draws from potential commitments, which means that one adopts a set of commitments from all the possible ways one could form commitments from one's complete set of motivational dispositions. So, taking a set of commitments from the field of potential commitments, one aims to do so in a way which achieves the greatest systematization – one aims to adopt the set with the greatest systematicity. Of course, since it is given that one is deliberating about what to do in a particular situation, one must aim to systematize with a principle that directs one to act in some way *in that situation*. The systematization must include something that answers the question, "What should I do in such a situation?" (or "this situation").

To arrive at a set of systematized commitments there are certain formal relations which must hold between the commitments in the set. The relations include compatibility, implication, and system unity. Compatibility is just the relation of having no contradiction or conflict between, or within a group of commitments. Implication means logical implication, although perhaps we should include something less formal, like a relation of "support." Maybe there are structural relations of one commitment supporting another which go beyond strict logical implication. In any case, we can approximate various relations of logical implication by considering looser relations of support. I say more about this later in this section. The other relation that is clearly a structural relation of affinity is system unity. System unity is in some ways, perhaps, the most important of these relations. This is the relation of being connected by implication or supporting relations overall. A group of commitments has system unity to the extent that the commitments mutually support each other and are tied together by relations of implication. If a set of commitments has two groups, each of which is thoroughly tied together by various implication relations unto itself, but which are completely isolated from each other in terms of implication and support, then that bifurcation is a lack of system unity. I discuss further how all of these relations are to be assessed when considered in the context of deliberation later on.

The deliberation as systematization model has two principle aspects. There is (a) what is formally required in a strict sense, and (b) there is what must be done from the perspective of an actual person in deliberation. The deliberation of an actual person can be further divided into (b_1) what the individual needs to accomplish in a formal sense, which must follow from what is strictly required for rational deliberation as such (that is,

(a)), and (b_2) how a person realizes such formal requirements in that person's actual experience. This gives us three categories for exploration: (a) what is strictly implied, (b_1) what a real person must formally accomplish, and (b_2) the phenomenology of deliberation. I deal with the first two categories in what follows and then I deal with the third category, phenomenology, in the next subsection.

What is strictly implied by my argument for systematization has really already been stated and, in a way, is rather simple. One has a field of potential commitments based on one's set of motivations that can be adopted as one's set of actual commitments. In rational deliberation, one aims to adopt a set of commitments with the greatest systematicity. Since one must draw from one's potential commitments and one must systematize with respect to *them*, strictly speaking, one is compelled to adopt a full system of commitments every time one makes a choice. Every time a choice is made, the whole system is up for revision. That is, every time one deliberates rationally and works toward a principle of action one must adopt anew one's entire set of commitments, where this set should be as systematized as possible.

This may seem to be an extreme conclusion (or at least a little funny) at first, but this turns out not to be so extreme when we look at how people must make decisions (which I take up below), and this also helps to make sense of something entirely intuitive. I take it as intuitively obvious that no commitment is absolutely binding in isolation, at least if there is nothing special about the particular content of the commitment which might make it binding, and it is intuitively well-supported that commitments are not absolutely binding independent of their content.[12] Whenever I deliberate, I can

[12] Of course, some philosophers believe that certain commitments are absolutely binding because of their particular content. Korsgaard, for example, believes that one is absolutely bound to the commitment to

potentially change my mind about any practical consideration (anything I value or take to be reason-giving) which arises in my deliberation. No commitment is absolute in isolation.

Commitments *are* binding in the sense that one is bound to them until one does in fact change them – this is what makes them *commitments*. Further, commitments are binding in the sense that a change in one may require a change in others, and there may be various options which are not open without changing some of one's commitments. In a more everyday sense of 'commitment,' there may be some things which one is bound to until one changes it, but, for some reason, one cannot change it. For example, we ordinarily think that a person is bound to a promise that has been sincerely made until the promisor changes his state of tying himself to what was promised *and* that the promisor cannot change that state without the consent of the promisee. This is a commitment in a fuller sense than I am using, although, of course, I do provide a sense in which one cannot freely change some of one's commitments – the tie to systematization.

In isolation, then, one's commitments are only relatively binding. They are binding relative to one's capacity to make necessary accompanying changes. However, it remains the case that one is bound to systematization. So, commitments can be binding as being part of the set with the greatest systematicity. Overall systematicity is not relatively, but absolutely binding. This implies that commitments that are tied to overall systematicity are themselves absolutely binding when considered as part of that commitment package. A commitment may be part of the determinate, unique set of

respect humanity, because one must value humanity in order to value anything at all Christine Korsgaard and others, *The Sources of Normativity*, ed. Onora O'Neill (Cambridge: Cambridge University Press, 1996).. This sort of view is what Sharon Street calls "substantive consructivism" Sharon Street, "Constructivism about Reasons" In *Oxford Studies in Metaethics: Volume III*, ed. Russ Shafer-Landau (Oxford: Oxford University Press, 2008)..

commitments with greatest overall systematicity, or it could be included in all sets of commitments that have acceptable systematicity. Either of these scenarios would make the commitment in question absolutely binding when considered globally.

So, in sum, what strictly follows from the argument for systematization is the following. One must draw from one's field of potential commitments by adopting a set of commitments with the greatest possible systematicity every time one makes a rational decision. No commitment is absolutely binding by itself, but, because one aims at overall systematization, commitments are binding as part of a package of commitments which jointly reinforce each other through systematic relations. The set of commitments one adopts through a decision can always (and in a sense must) be changed, but the systematic relations between commitments guide the way that individual commitments can be switched and altered.

Although, strictly speaking, one must adopt one's complete set of commitments anew every time one deliberates, for actual human beings at least, it is not feasible to consider complete sets of commitments and compare them with all possible alternatives by the standard of systematicity. People are just not capable of thinking that way. This does not, however, make it impossible for people to rationally deliberate (in the sense I have defined and argued for). Instead, it just constrains the way we do it. Instead of considering and adopting commitments in such a global way, we must do it in a more piecemeal fashion. We must consider commitments on an individual basis and see how they fit in with what we can discern to be relevant to the situation.

When I make a decision about what to do, I do not really need to consider revising every normative judgment I might potentially accept. All I need to consider is

what is relevant to the situation I am deliberating about. What makes something relevant is whether it has import for what I do in the given situation. A consideration has import for what I do in a given situation just in case (1) it is connected to a principle of action which applies to the situation by systematic relations, and (2) it is not ruled out by commitments which (relatively speaking) cannot be changed to accommodate the consideration because of their position within a systematic scheme. The first condition captures both things that support a consideration and things that count against it, because principles that conflict with something that applies to the situation must also directly or indirectly apply to the situation. For example, if I deliberate over whether to go get a book from the library today or not, I might consider a principle that I go after I finish reading some paper, in a situation in which I will not finish reading today. Going with this principle would be a decision not to go get the book today, which might be supported by other principles concerning my purpose for getting the book and how that relates to other things I have going on. Any consideration which were in conflict with this principle would also, either directly or indirectly, apply to the situation. A principle that challenged the relationship between my purpose for getting the book and finishing reading the paper, where the relationship supported waiting until I finished reading, would have application itself for what to do, or would have systematic relations with principles that had application for what to do.

The second condition for relevance, that the consideration not be ruled out by an "overriding" commitment derives from what is ultimately at stake in assessing one's considered commitments. One is trying to adjust those commitments that can be adjusted in a way which improves or maintains the systematicity of one's complete set of

commitments. Any consideration which only pertains to adjustments which would decrease overall systematicity cannot lead to the right choice and so is irrelevant for deliberation. (That is, an adjustment that decreases overall systematicity cannot lead to the right choice unless *every* decision in the situation will decrease overall systematicity, in which case, the issue becomes one of finding the smallest loss.) How some commitments can have the status of "overriding" others is something I address in more detail below.

In deliberating in this way, it is certainly possible that I might miss something that is relevant, but there is nothing unintuitive about that – of course we are capable of missing relevant considerations. In fact, any theory that did not leave room for an agent to miss relevant considerations would have a *prima facie* serious problem. While my theory implies that there will be many considerations that are relevant to deliberation, it is still reasonable to think that one can get things right without explicitly considering *every* relevant possibility. I take this to be a fairly standard feature of any serious theory of practical rationality.

As a starting point for deliberation which takes the form of considering commitments on an individual basis one must work with one's *actual* commitments. So far, I have only described potential commitments and commitments as they are adopted with each decision. There are, nonetheless, standing commitments in a certain sense, because people must deliberate piecemeal. Since one cannot throw out and re-adopt a complete set of commitments every time one deliberates, one is left with standing, actual commitments of two kinds: historical and dispositional. One has historical commitments in the sense that one has adopted various practical principles by making decisions, and

one accumulates these principles as standing commitments unless one changes or rejects them later on. Perhaps one can shed historical commitments without explicitly thinking about and rejecting them, but they stand to the extent that they continue to shape the way one makes decisions. If one operates as though one still accepted them, then they continue to stand. So, if I decided yesterday not to go to a certain bagel shop anymore, and today I still operate as though I don't go there by not giving the place my consideration, then my commitment to not going to that shop continues to stand. There may be some gray area with regard to which commitments one still has and which one has dropped, but that does not diminish the fact that people have many standing historical commitments.

The tacit acceptance of historical commitments is supported by the second group of actual commitments, dispositional commitments. Dispositional commitments are principles which one would adopt if various things were consciously considered – that is, if the principle were considered *or* if other principles would be adopted when considered which *implied* the principle in question.[13] If a person would adopt a principle based on normative thinking that brought the principle into view, explicitly or implicitly, then that principle is an actual commitment for the person. So, if there is some other bagel shop I am familiar with which has features that, if thoughtfully considered, would lead me to adopt the principle that I always go there when I am in the area and hungry for a bagel, then I have a dispositional commitment to regularly going to that bagel shop.

[13] There is also the issue of different brands of normativity implying each other or being equivalent in some way. For example, while my theory is worked out in terms of principles which I have argued are necessary for rational deliberation, other kinds of normativity like practical reasons or values may imply practical principles when taken in conjunction with claims about the fundamental nature of principles for practical rationality. This means that a person who values certain things, or accepts certain practical reasons, thereby implicitly accepts corresponding practical principles or commitments. I will say more about possible equivalence relations between principles and other normative objects in III.B.

These actual commitments are in no way guaranteed to be well-systematized. They form a complicated web that may be extremely problematic in terms of systematization. There may be outright contradictions between them, there may be commitments which are not connected systematically with anything else, and there are surely many conflicts that are not outright contradictions between principles, but which are problematic nevertheless. These problems for systematization are what the agent is formally compelled to address in deliberation.

So, we are rationally bound to deliberate by adopting a principle of action in such a way that it is as systematized with our actual standing commitments as possible while adjusting the problems for those standing commitments as needed. Problems for our commitments fall into two main categories: conflict and disunity. The problem with conflicting principles is rather clear. Principles can conflict either formally or in application. The purest and most straightforward type of formal conflict is when principles taken in conjunction involve a contradiction. This is perhaps the greatest violation of systematization and must be resolved by changing enough of the principles to remove the contradiction. There may be other types of formal conflict between principles that, for example, depend on analytic content of principles in some way that is not the same as a purely logical contradiction. I leave this possibility open.

The second category of conflict between principles arises in their application. I may, for example, have a principle to drink water whenever I am thirsty, although this is not always possible, so this principle may be said to have conflict in application all by itself. A further kind of conflict in application comes from principles which cannot be jointly applied, either because this is always impossible or because it happens to be

impossible in a particular case. An example of the first kind would be principles to always walk the entire distance between two places, and to always bike the entire distance between the same two places. It is always impossible to jointly apply these principles that are supposed to both be universal in nature. An example of the second kind would be principles to always buy a coffee when passing a certain shop, and to always buy a sandwich when passing the same shop. I may pass the shop at a time when I do not have enough money for both, which creates a conflict in application that only arises because of the particular circumstances.

All of these kinds of conflict, and any others which I did not mention, require resolution of some kind through deliberation. They require that various adjustments be made in one's set of commitments which resolve the conflicts. At least, adjustments must be made which resolve the conflicts that pertain to the decision under consideration. This resolution of conflicting principles I call "harmonization." In the process of harmonization, at least some principles must be given up in order to resolve the conflict. Which principles must be given up, however, is not given by the mere fact that conflict arises. For example, if one holds a principle of not cutting in line, p_1, but, at a time when one is in some government office where there is a long line of people waiting already, one forms a principle of getting through the government office as fast as possible as determined by some subjective risk analysis, p_2; then one has two principles which are presently in conflict (assuming there is a way to get an advantage by cutting in line). One aims at harmonizing them, which means resolving the conflict. The mere fact that there is conflict does not require that p_2, the new principle, be dropped. Harmonization only

requires an attempt at making the conflict go away. One way of doing this would be to revise p_1, another to revise p_2.

Next to conflict, the other major category of problems for sytematicity is disunity. Principles which are unconnected to one's overall set of commitments are not systematized with that set, and so they present a problem for systematicity. Similarly, if groups of commitments are internally systematized but are not systematized with each other, this also presents a problem for systematicity. Intuitively, this is a problem because, in this respect, the agent is not unified and, with respect to these separations, is not capable of deciding on the basis of a single overall rational decision-procedure. More formally, this is simply a lack of systematization. When one seeks to make a decision in a given situation, the creation of the principle of action for the situation must be aimed at systematization with one's commitments or the decision does not come from one's deliberation at all. Furthermore, one needs to systematize the commitments with which one supports the principle of action, and bring them together if they are not connected by systematic relations. Other than the need to systematize the new principle of action, this issue will mainly arise in conjunction with an issue of conflict. If I am deciding whether to keep some money my employer mistakenly paid me, I might be working with a group of commitments which imply that I should not keep the money and another group of commitments which imply that I should keep the money. If these two groups of commitments are not connected to each other by any systematic relations, then part of the problem I have is to create unity between these commitments (while the other part of the problem is to resolve the conflict).

I call the process of systematizing and ordering one's commitments by connecting them with subsuming principles "unification." In the process of unification one tries to connect hitherto unconnected commitments by bringing them under some commitment that can be taken as antecedent to them. This is not to say that one discovers root commitments that were always behind the considered commitments logically or psychologically, but rather that one *adopts* a common or unifying commitment which can tie together disparate aims, purposes and values. In order to do this, one will have to conform one's actual commitments so that they can be brought together by a common principle.

The problem with what has been said so far is that, on the one hand, we must work to systematize on the basis of individual commitments because of our limitations as practical reasoners; whereas, on the other hand, the problems that arise for commitments on an individual basis do not seem to bring a solution that can be applied without comparing overall systematicity. Initially, there does not seem to be a way of doing piecemeal practical deliberation that truly aims at overall systematization. This might seem to make it impossible for us to meet the requirements of rational deliberation as they have been laid out.

However, while, formally speaking, one must adopt a whole system with each choice, there are traits of individual commitments, deriving from general systematicity, which make harmonizing and unifying commitments on an individual basis capable of determinate correctness. Specifically, some commitments will "do more work" than others in systematizing one's set of commitments, either potentially or actually. *Doing more work* just means being more central in the systematic web. In other words, a

commitment A does more work than a commitment B just in case A has more relations of systematicity with other commitments in the set, including implying others and being implied by them. The work done is the work of systematizing, which is done by being in systematic relations.

If A does more work than B, then if the decision comes up whether to keep A or B, then, *ceteris paribus* (a clause to be discharged immediately below), A should be kept at the expense of B. This follows from the basic aim of systematization: whichever scheme has greater systematicity is a better choice. When considering two competing principles in isolation, the principle which does more work will contribute to the scheme with greater systematicity, and so is a better choice than the principle which does less work.

Of course, there can be cases where, strictly speaking, A does more work than B, but of all the systematizing schemes or sets of commitments, the one that has the greatest systematicity includes B but not A. A real example might be something like the following. My principle to help a person needing directions does more work than my principle to hurry to the subway. However, under the circumstances, my best choice is to hurry to the subway and leave the confused person, because I need to get to the box office to buy some tickets before it closes. While this complication does reveal some of the complexity of deliberation, this does not overthrow the general standard of authority based on systematic work. It is not as though the principle to hurry to the subway actually does more work in the situation, but rather, there are other principles which jointly do more work than the principle to give directions that apply in this situation. The

greater authority comes from other principles which, under the circumstances, support hurrying to the subway.

With this standard of authority, my account of what a real person must formally accomplish in rational deliberation is complete, at least as an outline with all of the major issues addressed. While there are certainly more details that could be explored and filled in, such as a fuller account of the relative systematic work done by individual principles, I have sketched out how such details are to be filled in and where they are to come from. These requirements take into account some of the key limitations of a human's ability to systematize potential commitments. At the same time, this account shows how a human can pursue the rationally required aim of deliberation and have a standard for correct deliberation, even with the inherent limitations of a human being. What remains now is to investigate the relationship between these formal requirements and one's experience and intuitions as a deliberator.

B. The phenomenology of deliberation

The third and final issue to cover in my account of rational deliberation is the phenomenology of deliberation – how our deliberative experience and intuitions about rational deliberation relate to the formal requirements I have argued for. In providing this account of the phenomenology of deliberation, I not attempt to give a definite and final version of how deliberative experience relates to the formal requirements, but instead, I simply try to show that the formal requirements can be reconciled with our experience and intuitions.

There are two primary problems for the phenomenology of rational deliberation. The first problem is that, (a) according to the requirements, there are very specific forms our considerations and decisions have to take which people may seem, in general, not to

satisfy. Specifically, I have argued that decisions must come in the form of practical principles and that deliberation must come in the form of manipulating more principles. It may seem, according to our intuitions and internal observations, that for many of us this is simply not the form that our decisions and deliberations take. Instead, our deliberation may seem to take the form of the manipulation of simple, non-practical-principle reasons or statements of value, for example. Maybe my deliberation takes the form: "Under the circumstances, being out of milk is a good reason for me to go to the store. I value having milk in the morning more than I value not having to go out the night before. Therefore, I will go to the store." If this is the form my deliberation takes, it seems that I have not manipulated any principles to come to my practical conclusion, and my practical conclusion does not take the form of a principle. What is required to resolve this disparity is an argument for various equivalence or implication theses for these different forms of normative reasoning.

The second major problem for deliberative phenomenology (b) concerns the status of our intuitions about rational deliberation and moral reasoning, as we know it from public discourse. The way we are used to engaging in moral reasoning, at least as a public exercise, is by arguing that some policies or actions make more sense with our general views about how we should act and how society should be run. At least in philosophical applied ethics, we appeal to theoretical principles from moral theory and argue that these principles are the most correct or the most applicable for the case at hand. Perhaps in less philosophical discussion, the same thing is done but with less care for rigorous definitions and arguments.

The problem that this raises for the view I am arguing for is how these common forms of moral reasoning are supposed to be understood in terms of systematization. If common moral reasoning cannot be understood in terms of systematization, then I have to admit that a *modus tollens* argument that there must be a problem with the systematization theory seems more compelling than a *modus ponens* argument that there must be something wrong with standard moral reasoning. Hence, I need to explain how standard moral reasoning fits with the systematization account of rational deliberation.

(a) In order to handle the first problem, I more or less need to provide suitable equivalence or implication hypotheses about the language of deliberation. As is customary, I carry this out on the assumption that this is more or less a question of interpreting sentences in natural languages which are the vehicles for private or public deliberation. To the extent that it is possible, I would like to leave open the possibility of interpreting deliberation, at least of the private kind, in a way which does not wholly or even at all use natural language as a vehicle. For example, one might come up with a theory of how a deliberator can "summarize" the propositions typically manipulated in deliberation through some other means than through standard explicit sentences. This is obviously a large and deep issue in the philosophy of mind and language, which I do not pretend to be able to explore with any notable insight. Hence, I deal with the first problem by considering a few standard English linguistic devices which I take to be typical of how deliberation is done (and to have suitable equivalents in other languages), and to express modes of thinking and doing that are supposed to be satisfactory for rational deliberation. Specifically, I argue that reasonable theories of implication are available for practical conclusions of the form, "I will do X," (and similar expressions)

implying "I do X on condition Y" (the form of a principle of action); and that reasonable theories are available for other normative expressions, as used in deliberation, being equivalent to expressions mentioning only the elements of my theory of deliberation. In this way, I argue that many of the forms of deliberation which are typically taken to be, and out intuitions tells are rational deliberation *do* satisfy the specific forms of deliberation required by my arguments. So, the objection, "but I don't think that way when I deliberate, and I don't know anyone who does think that way when he or she deliberates, so there must be something wrong," is misplaced.

At least as an important example, practical conclusions can take the form, "I will do X." This captures the determinative content of a practical conclusion. I take it that practical conclusions necessarily are accompanied by the corresponding act, otherwise they are not truly practical conclusions, but in any case this does not distinguish the two forms, "I will do X" (with background reasoning) and "I do X on condition Y." So, the issue is whether "I will do X" with appropriate background reasoning can be taken to imply or be equivalent to "I do X on condition Y" where Y comes from the background reasoning for the practical conclusion. I claim that it is entirely plausible that if an agent A really does embody authorship in a decision – that is, A exhibits agent-causation by way of rational deliberation – then what A means by things like "I'll do X now" or just "I'll do X" is "I do X in this situation," where "this situation" is understood in terms of whatever A considered in making the decision. This is a perfectly reasonable interpretation for exactly the reasons I gave in Chapter 2 for the claim that making decisions in the form of a principle of action is required for rational decision-making. If "I will do X" does not mean that A does X universally in the situation as conceived in

A's deliberative reasoning, then A's deliberative reasoning is not the cause (even in a non-exclusive sense – never mind the reasoning being the only or the true cause) of the act, and so the act does not amount to an action by A. A states the decision in terms of what A will do now, only to emphasize the tentative, revisable nature of the decision – *not* to make it a "singular edict."[14] Perhaps forming the conclusion in terms of "I will" not only makes it feel more revisable, but also helps to reinforce executing the decision in a situation in which the agent does not feel absolutely certain that the right choice has been made. The principle of action is implied, although it is not emphasized. So, if the agent really does do what is necessary to perform a true action, then the agent does mean or intend to do some act X on condition Y, universally, and so accepts the principle of action, "I do X on condition Y."

The next point of equivalence to address is how to handle standard appeals to what a person values and what counts as a reason, in the most common practical sense. If I can show that these appeals can be reasonably interpreted in terms of principles and systematization, then I will have put my theory in relatively uncontroversial territory with regard to how we intuitively understand deliberation. The first step in this argument is to point out that there are already theories available for how values can be understood in terms of reasons, notably T.M. Scanlon's "buck-passing" account of value (Scanlon 1998, 95-100). According to this theory, "[t]o value something is to take oneself to have reasons for holding certain positive attitudes toward it and for acting in certain ways in regard to it," and, "[t]o claim that something is valu*able* (or that it is "of value") is to claim that others also have reason to value it, as you do" (Scanlon 1998, 95, author's italics). The attitudes and ways of acting a person has reason for in valuing vary with the

[14] See the discussion of the possibility of "singular edicts" in Chapter 2, Subsection IV.C.

cases, as Scanlon states, but since this is not a free-standing account of value, it is not necessary to provide rules for this. The point is just that value can be understood in terms of reasons.

If values can be interpreted in terms of reasons, then all I need to do is provide a plausible interpretation of reasons in terms of principles and systematization. Since the issue here only concerns reasons as used in deliberation, I only consider reasons for particular agents. As a first attempt, consider the suggestion that reasons can be understood in terms of revisable principles. According to this suggestion, "being out of milk is a reason to go to the store" implies "I go to the store, when I'm out of milk and I have no overriding principle not to." Principles of action (commitments), as I have made very clear, are always revisable. Having a principle does not settle the issue of what one should do. If I have a principle to go to the store on condition Y, I can change it, but, in a sense, I have a *reason* to go to the store on condition Y. This is because I have such a principle, it directs me to act so, and it forms part of my systematization of commitments. The conceptual role that 'reason' plays in deliberative discourse is *counts in favor of* (Scanlon 1998, 17). The conditions for having a commitment which I just listed satisfy this criterion – they jointly *count in favor of* going to the store. To this extent, having a certain principle is a valid interpretation of having a reason. It may be unsatisfactory though, because this interpretation would make it difficult to smoothly transition to third-party ascriptions of reasons. If you want to say of me that I have a reason to go to the store to get milk, we should be able to make sense out of this without implying that I accept a principle about going to the store. Even if it has to be the case that my having a reason must come from the systematization of my commitments (indeed, this is a central

claim of my theory), it seems that it should be possible for me to have reasons which do not *directly* derive from some particular accepted principle of mine. Roughly speaking, we should be able to understand reasons in a way that allows for me to have reasons prior to figuring out that I have a reason.

Now consider the modified suggestion that reasons can be understood as summaries of systematization. According to this suggestion, "being out of milk is a reason to go to the store" implies that my set of actual commitments is such that principles concerning being out of milk and principles concerning going to the store (going to the store in the sense in question) do more to support than to conflict with the principle "I go to the store when I'm out of milk." This account has the advantage that it captures the "counting in favor of" by the balance of systematization concerning going to the store and being out of milk, while leaving open the implications of other kinds of principles that might apply to the situation. For example, the situation might be one in which the agent's leg hurts and so the agent "has a reason" not to go the store. This additional fact would make principles concerning having a hurt leg apply to the situation and allow for the possibility of such principles doing more to support than to conflict with a principle, "I don't go to the store for milk when my leg is hurt." So, the defeasibility of reasons would be captured by this interpretation of reasons as summaries of systematization.[15] Furthermore, this version clearly does not have the problem that my first suggestion had, namely, to make it difficult to see how an agent can have a reason based on the agent's commitments and systematization, which does not directly derive from an actual commitment. Here, reasons spring straight from the overall

[15] For more on the importance of the defeasible nature of reasons, see Jonathan Dancy, *Ethics without Principles* (Oxford, Clarendon Press: Oxford University Press, 2004).

systematization of the agent's commitments and do not depend on the existence of any particular commitment. I take the satisfaction of these traits of practical reasons to show that practical reasons can be understood in terms of principles and systematization, and so my theory of rational deliberation does not fly in the face of accepted and intuitive ways of thinking about rational deliberation.

(b) In order to handle the second problem, how commonly accepted and intuitive forms of moral reasoning, which I will call "standard" moral reasoning, are supposed to be understood in terms of systematization, I need to argue that standard moral reasoning performs the work of systematization. The problem is that we have various notions of what moral reasoning and explicit practical deliberation are, but, according to my theory, the work of deliberation is to develop the systematicity of an individual's set of commitments. It may not be obvious how standard moral reasoning does this, which means that it is not obvious how standard moral reasoning is at all relevant to figuring out what one should do. If I can show that commonly accepted types of moral reasoning help us to systematize our commitments, and thereby help us determine what principle of action to adopt, then I will have shown that standard moral reasoning does the work of rational deliberation – it helps us figure out what we should do.

In order to provide such an argument, I need to talk of commonly and intuitively accepted moral reasoning very bluntly and crudely, and I need to make rather hasty generalizations about how it typically goes. However, I believe that these blunt, crude characterizations are roughly on the mark and that the somewhat sketchy argument I provide could easily be expanded and refined to accommodate a more refined account of

standard moral reasoning. Hence, for my present purposes, the argument for moral reasoning as systematization that follows should be sufficient.

In accordance with what I said about the formal conditions for rational deliberation, I need to show that standard moral reasoning harmonizes and unifies in accordance with the standard of authority. Now, standard moral reasoning often works to draw out conflicts between principles of action and principles of moral theory (utilitarianism, Kantian imperatives, principles of distributive justice, various supposed rights, standards of virtue, etc.) which the participants are supposed to accept. If the participants do not accept the proposed principles of moral theory then these moral theoretic doctrines are argued for on the grounds of other principles of moral theory or general practical principles, which are sometimes drawn out through intuitions and thought experiments. So, the basic form of standard moral reasoning is to argue with moral theoretic principles to practical conclusions.

Doing just this, arguing for and with moral theoretic principles, is a sweeping way to address harmonization and unification from the basis of authority, since the whole point of arguing this way is that the theoretical principles which are brought out to direct one's actions or policies are supposed to have more grip on the participants than any principles of action which would conflict with them. They are supposed to have more grip in the sense that they make more coherent all of the principles of action or practical reasons we accept. Moral theoretic principles are by their vary nature subsuming principles which are supposed to imply "the right" principles of action, which means that their job with respect to systematization is to imply or explain other principles of action. Therefore, the work of picking out moral theoretic principles which explain our accepted

principles of action in balance with adjusting our principles of action to accommodate overall highly "explanatory" moral theoretic principles is exactly the work of systematization. Hence, standard moral reasoning, as I have here interpreted it, does precisely the work of systematization. Moreover, unphilosophical standard moral reasoning, insofar as it is like philosophical standard moral reasoning with less rigor, also does the work of systematization, although it may do so more roughly.

IV. Resolution of the problem of practical normativity

In Chapter 1, I argued that the main issue for explaining the possibility of morality and reconciling objective morality with a naturalistic worldview consisted in substantiating what I called "the reality of normativity." The reality of normativity is the normative aspect of morality, as distinguished from the aspects of morality that concern its content. While for moral obligation, normativity and content may be the whole picture, this is not clearly the case for morality all together. There may be more to morality than a set of obligations. But, as I argued in part in Chapter 1, and as I argue more thoroughly in the remaining chapters, an account of moral obligation of the right kind can be sufficient for an account of objective morality. This makes the normative aspect of moral obligation not only a key point of contention for naturalist moral skepticism, but also a potential building block for a general account of morality.

In the last chapter and this one, I have attempted to supply an explanation of practical normativity, which can be used to set up an account of morality later on. This explanation of normativity starts with constitutivism, the view that practical normativity is grounded in the constitutive aims of agency. I have argued that an agent performing an action, the fundamental notion in practical philosophy, happens through reflective

endorsement based on principles of action. This embodies two ideas that have enjoyed much currency in practical philosophy. Reflective endorsement centers around the idea that agency is a matter of *unified*, *reason-based* decisions. The principle requirement goes back to the Kantian idea that the decisions of an agent must be made in *universal form*. I have argued that both of these ideas get at something essential to agent-causation.

To further press the constitutivist claim, I first sketched a general argument that an agent must aim at self-unification. In this chapter I replaced that argument with a more formal argument that reflective endorsement must happen by drawing from potential commitments, and that drawing from potential commitments in rational deliberation must be done by aiming at systematizing one's commitments. These arguments lead to the conclusion that the constitutive aim of agency or, more precisely, of deliberation is systematization of one's commitments. This allows me to state my version of constitutivism as the claim that practical normativity is grounded in the aim at systematization of one's commitments.

My explanation of practical normativity is constructed from the account of practical rationality – that rational deliberation aims at systematization of commitments – and four theses that link this account with normative standards for action: Deontic Kantianism, Attribution, Transitivity, and Assessability. These four theses allow me to conclude from my account of practical rationality that if an act is normatively assessable, then it is right or wrong in accordance with its compliance with the aim at systematization. This is an explanation of normativity.

This explanation of practical normativity supplies what I argued in Chapter 1 was the most important step in substantiating objective morality. This may not be enough by

itself to talk of "moral obligation," but it goes a long way in bringing us to moral obligation. The account of normativity gives us a framework within which moral obligations can take shape and find their footing. Moral obligation only makes sense if the normative aspect of it makes sense, and it is this normative aspect which has been attacked most heavily by philosophers skeptical of moral obligation. Providing a format for substantive, content-providing theories to take hold is what a first-personal normative explanation does.

However, the way I have explained normativity leaves right and wrong action dependent on what a person's contingent motivations are. This may seem to create an insurmountable problem for constructing a theory of morality with my account of normativity. In the remaining chapters I argue that in fact this is not an insurmountable problem and that a theory of morality can be based on such an account of normativity. This will involve arguing that a minimal conception of morality can be built out of this account of right and wrong action. To make such an argument, I set out some requirements for the content of moral obligation and the nature of moral discourse and argue that these requirements can be satisfied to a sufficient degree, even by a theory which leaves normative content dependent on individuals' contingent motivations. In the next chapter, I argue that the form of moral obligation set out by Stephen Darwall can be satisfied by my theory (Darwall 2006;Darwall 2007, 111-132;Darwall 2007, 52-69). This involves arguing that my theory of practical normativity can handle the second-personal foundation of moral obligations, which, according to Darwall, must come from second-personal accountability relations. I explore this issue in terms of a puzzle about practical normativity and moral obligation.

Thereafter, in Chapter 5, I argue for a few requirements concerning the content of moral obligations and the role of moral discourse which a theory of morality should meet. I then attempt to show that my theory can handle these requirements. Finally, I argue that satisfying these requirements is good enough for an account of right and wrong action to count as a theory of morality. If these arguments are successful, then I will have completed my explanation of the possibility, and the actuality of morality.

Chapter 4
Moral Obligation: Normativity and Social Accountability

I. Introduction to Part III

In Part II, I laid out and argued for a constitutivist theory of practical normativity, which is by its nature highly explanatory. I used explanatory arguments to attempt to show that the guiding aim of deliberation is systematization of one's commitments, and that this guiding aim dictates what right action is. My conclusion from Part II was the proposition: *If an act is normatively assessable, then it is right or wrong in accordance with its compliance with the aim at systematization*. Further, I explained that normative assessability is a question of the ownership or attributability of action. I argued that *an act is attributable to a given agent if and only if it is an action that agent is the author of or it is behavior caused by habituation, training or preemption by (an) action(s) that agent is the author of*. This last thesis, which I called the Attribution Thesis, has the virtue of being able to handle the issue of how deliberative, practical reason is the source of agent-causation (rather than, for example, anything to do with metaphysical freedom of the will), *and* the issue of how unreflective behavior of various kinds can also be cases of agent-causation or acts attributable to an agent.

The systematization of abstract, flexible commitments as the core of practical reason and the source of practical normativity may seem excessively complicated and remote. However, I have provided careful, detailed and, I hope, rather rigorous arguments for my theory. Just as important, the theory has the virtues of being highly

explanatory and naturalistic, as well as that of tying together solutions to various problems in practical philosophy. These issues include the problem of agent-causation (how an agent can be the cause of its behavior and own the action), the mystery of where normative force comes from, the question of what common standard practical reasons all share with one another, and in this chapter I will try to show that my theory gives an answer to questions about the relationship between practical reason and morality, and a resolution to a standing tension between social and individualistic theories of morality. This is the second major problem I referred to in the Introduction – that of explaining the general authority of social accountability relations in terms of a pre-explained general scheme of practical normativity.

To the end of making a definitive tie between practical reason and morality, I argue for a theory of morality that is grounded in my theory of practical normativity. As it stands, I do not think the theory of practical normativity I presented in Part II can be considered a moral theory. It does not obviously bear the characteristics of any established conception of morality: I have not drawn any specific practical consequences from the theory, and I have not provided a connection to moral virtues, cooperative living, or tied it to an analysis of moral obligation. I do not believe any of these features is, by itself, essential to morality, but I take it that a theory of morality should include things that at least amount to a family resemblance to this group of features.

In this chapter, I cover much of the ground of providing a moral theory based on my theory of practical normativity. I address the issue of social accountability, which is a key ingredient to morality. Specifically, I look at Stephen Darwall's recent theory of morality based in social relations, which he calls "second-personal reasons" (Darwall

2006; Darwall 2007a, 111-132; Darwall 2007b, 52-69). I analyze Darwall's view, which I think partly correct, and show how we can drop the part of it that does not work and replace that with my theory. That is, I analyze social accountability in terms of second-personal relations and then explain how there can be a system of accountability-based obligations for an arbitrary agent in my normative framework. This results in a general moral theory with some very interesting characteristics, among which is that the phenomenon of moral obligation is empirically contingent and that the normative supremacy of moral obligation is merely general among rational agents, not universal.

The path to a theory of morality is to set up a theory of moral obligation, which I take to be sufficient as a theory of morality, even if it is not necessary. Moral obligation may not be all that there is to morality, conceptually, but it is enough that a theory of moral obligation can be a theory of morality.

However, at the end of the chapter, I also add some remarks about how a kind of a theory of virtue also falls out of my theory of practical normativity, so that the picture we get includes more than what is strictly contained in the account of obligation.

The way that I come to a theory of obligation is by looking at the essential characteristics of obligation and how they can fit with my theory. Roughly, the key characteristics of obligation I explore are the following. (1) That obligations arise from a demand that an individual or the moral community at large is in a position to make of the agent. So if Charles has a moral obligation to be considerate to his neighbors, then this obligation must arise from demands that the neighbors or the moral community at large are in an authoritative position to make of Charles. Whether or not they need to in fact make the demand in some way is a further issue, that may depend on the particulars of

the case. (2) That moral obligations enjoy normative supremacy – normative authority which overrides other normative or motivational pressures. Charles's moral obligation to be considerate to his neighbors has normative supremacy if any other reasons Charles has to not be considerate to his neighbors (like reasons of self-interest) are outweighed by his accountability to his neighbors to be considerate of them. David Brink calls this the "supremacy thesis" (Brink 1997, p. 255). Stephen Darwall provides powerful arguments for the first condition, which I subscribe to. Darwall accepts the second condition as well, although ultimately I use this against him to refute part of his project and make room for my own theory.

II. Morality as a social phenomenon

A. The issue of social relations

A general question for metaethics is whether and to what extent morality or normativity is fundamentally social. The alternative is that it is "individualistic," in the sense that it could apply to someone in the absence of any social relations. Some hold that morality is entirely social – that it is a social construction in the sense that it is embodied by, or it arises from social conventions, perhaps of an unavoidable kind. For example, morality might be the result of social conventions which exist in order to provide a workable system of cooperative living and there might be a limited range of what works as a system of cooperative living with certain aims. Both Hobbes and Rawls held views like this. Hobbes, in fact, thought that certain practical rules could be directly derived from our social circumstances, although his theory had less of a social slant than most contractarian or contractualist views (Hobbes 1651, 1968). Rawls, whose theory is

very socially oriented, thought that our social aims constrain the available solutions to the problems of cooperative living (Rawls 1971). On the other side, some hold that morality is completely individualistic, in the sense that moral prescriptions are fundamentally no different than whatever prescriptions there are for rational agents or agents who are guided by norms in the absence of any social relations with others. Kant, for example, held such a view, since the categorical imperative, according to him, followed simply from being a rational agent (Kant 1785,1911). Moral prescriptions are just like other practical prescriptions for agents, Kant thought, because both are a matter of acting on practically coherent maxims. All deliberative decisions, for Kant, are equally moral decisions.

If we think of the issue as the question of what puts a normative matter in the domain of morality, the answer might be a certain tie to social relations. On the other hand, it might be just the degree of importance. This latter idea can be illustrated with the following examples. Which pencil I use to draw with is not very important (except perhaps in some bizarre circumstance) and maybe that is why it is, according to this suggestion, not a moral issue. Whether or not to kick a pigeon that has annoyed me is probably quite a bit more important, in an impersonal sense, and might qualify as a moral issue. Whether or not to poison my brother and take the family inheritance is a very important matter which might be why it would qualify as a moral matter. There will surely be difficulties for an importance-based interpretation of the moral domain, but it has a lot of potential.

At this point, I do not want to settle the question of what characteristics put a normative matter into the moral domain – whether it be social relations, importance,

suffering and happiness, or something else. Whatever the answer to that question is, the issue of how social relations are involved in morality, and what dependence morality has on social relations would remain. Further, it would remain open how it is that having a moral obligation would automatically imply that others have a right to hold the obligated party accountable and to enforce the obligation with sanctions and reactive attitudes. This issue, how morality depends on the social, and how social accountability is a part of morality, is central to how I construct a moral theory out of my normative theory from Part II.

B. Where constitutivism and systematization fit in

Constitutivism in general provides individualistic, internal reasons. It may seem at odds with a social take on morality. Constitutivism would seem not to provide a social theory of practical normativity or of morality. The only obvious way in which a constitutivist theory would be a social theory of morality would be if some social aim or motivation is supposed to be part of the constitutivist package. In other words, if something social were supposed to be constitutive of agency, action or deliberation, that version of constitutivism would be a kind of social theory of morality. Perhaps, for example, one might hold that forming agreements with the community, or as a community is constitutive of agency. Then the norms flowing from this aim would be of a fundamentally social nature. This is not my view and I do not take it to be very plausible, but it would provide a social theory of practical normativity.

My theory appears not to be social at all. It looks completely individualistic and at odds with any social approach to morality involving a foundation on the goals of cooperative living or social demands. Systematization of one's own commitments does

not imply anything about social relations, at least not in any obvious, uncontroversial way. This point could also be made in terms of my theory being so thoroughly internalist[16] as follows: Some moral theorists would only require that one have a motive insofar as one is rational, which is more of an internalist position than letting rationality and motivation be completely independent. My theory, however, makes motivation a part of the standards of rationality. Being practically rational is acting on well-organized motivations. Having motivations and organizing them seems to make everything happen on an individualistic basis, since these are not social activities. This makes my theory look irretrievably individualistic.

Also, if we look at the normative level, my theory seems to support a brand of deontology, which means that not even a consequentialist interpretation of social morality is available. Since, loosely speaking, my theory makes the agent's intentions what determines the rightness or wrongness of an action, it would seem to be a paradigmatically deontological theory. Although I do not want to make any final pronouncements here about what my theory implies about consequentialism, a consequentialist normative theory seems to be at odds with my explanatory theory of practical normativity, and would seem more suited to fit with a classical realist, externalist theory of normativity.[17]

---

[16] I discuss important varieties of internalism at length in Chapter 1, Section VI.

[17] In fact, the issue is a little more complicated than this, because what is normative for an individual depends on the content of that individual's motivational dispositions and commitments. This blurs the issue of where a normative theory is supposed to fit into the picture. As I argued in the last chapter, moral theories, on my view, do the work of systematization. (See Chapter 3, Section III.B.) This means that a moral or normative theory's job is to organize principles of action. If the best way to organize my principles of action is by focusing on the empirical consequences of my and others' actions, then a consequentialist normative theory will do the most work of systematizing my commitments, and would, therefore, be "correct" in my case. How we are to make sense of moral theories in the context of a public discussion is a topic I will address at length in the next chapter.

Setting the more metaethical distinctions aside for the moment, consequentialism would be a step closer to a social account of morality and right action, because the consequences for others would count directly toward the rightness of an action. On my system, we are a step removed from this, since it must first be part of the content of one's commitments that the way the agent treats others is of significance, or that some general variety of consequences (like pleasure and pain) is relevant or important.

To illustrate these different gradations, let us say that a general is deciding whether or not to bomb a village. On a social account of morality, like some contractarian theory, the question of how this relates to other people, possibly including the targeted villagers, is directly at issue. The general's social relations with the villagers, the bomber pilots, others in the military and the government, as well as his own people, are what determine whether it would be right, permissible or wrong for the general to order the village to be bombed, according to a contractarian, or other social theory of morality. According to a consequentialist theory, what is at stake is what consequences there would be from bombing or not bombing the village. On almost any serious version of consequentialism, this will include the bad (and perhaps good) consequences for the villagers, as well as some specified consequences for the same list of people specified above, and perhaps many more (perhaps all current and future generations of life on this planet and others). What is right for the general to do concerns a lot of the same facts, according to consequentialism as it does according to contractarianism. However, the contractarian gives social relations an immediate relevance that the consequentialist does not.

My theory seems to make social relations one step further removed from what determines the right course of action for the general. In order for the plight of the villagers, the bombers, the general's fellow citizens, or the rest of the world's people to matter at all in what is right for the general to do, the general's commitments must have some content that brings such factors into play. It could be that the general's commitments give authority to the consequences for all other people or for those who can expectably be affected by his actions, but it could also be (for all my formal theory has to offer) that the general only has commitments which concern his own well-being and how he will prosper or suffer from his actions. So, my theory has social relations even further removed from the field of relevance than a consequentialist theory. There is more of a chance that they wouldn't be relevant or that they would have less weight, than on a consequentialist moral theory, and especially a social moral theory like contractarianism.

C. What will follow

Stephen Darwall's recent work shows that moral obligations are fundamentally social in the sense that they must arise through second-personal relations and, consequently, they must possess a second-personal character (Darwall 2006; Darwall 2007a, 111-132; Darwall 2007b, 52-69). Darwall also argues that the second-personal aspect of moral obligation (and related concepts) cannot be reduced to first-personal relations or any normative relations that are not already implicitly second-personal. This is a powerful result which goes a long way in addressing the issue of what kind of dependence morality has on the social.

Since my theory seems to fall so sharply on the individualistic side of the issue, Darwall's arguments look like they signal a huge difficulty in putting together a theory of

morality out of my theory of practical normativity. It would seem that there is no room for me to assemble the needed elements of a moral theory, because the social aspect, that is, the second-personal aspect of morality does not seem to fit with my theory of normativity. Reconciliation looks to be out of reach.

However, I argue that second-personal relations of the kind Darwall argues for cannot, of themselves, generate a general system of moral obligations which have normative supremacy over other reason-giving phenomena for an arbitrary individual. In keeping with his arguments that the essence of morality is a certain kind of second-personal demand, Darwall tries to show that a foundation of unavoidable presuppositions provides the normative ground for a general system of practical norms. These norms flow from what he calls "second-personal reasons" which form the basis for a normative system upon which some kind of contractualism can rest. In short, Darwall tries to use a purely second-personal apparatus to ground morality. I argue that this attempted grounding of morality fails. Briefly, my argument draws on the requirement of normative supremacy – that moral prescriptions must override any other prescriptions – and I try to show that this requirement is unobtainable for the pure second-personal approach. This is because normative supremacy must be deliberative supremacy, which itself comes out of the first-person standpoint.

These arguments by themselves do not fix the problem, however. Showing that there is something wrong with a completely second-personal approach does not show that the way is open for a first-personal approach. Even if my objections hit their mark, I still need to show how my theory can handle the arguments that morality, at its core, requires second-personal relations. The way I get to such a conclusion is by first setting up a

general puzzle. I state more formally the tension between what Darwall shows successfully and what follows from my objections to his constructive arguments. I put this puzzle both in terms directly relevant for my project, and in terms that make the problem more general for metaethical theory. I then provide a solution to the puzzle which resolves the tension for my theory and lays the ground for a moral theory based on my metanormative theory of Part II. This solution also provides a model for a general solution to the general puzzle about the tension between the social and the individual, or, more accurately, between the second-person standpoint and the first-person standpoint in moral theory.

III. Darwall's arguments

A. Darwall's overall strategy

In *The Second-Person Standpoint*, Darwall pursues a two-part strategy. He first argues for a purely conceptual thesis connecting morality to the second-person standpoint, and then he goes for a more substantive thesis which is supposed to establish the authority of moral obligations. The first thesis Darwall sets up by arguing for a conceptual tie between responsibility or accountability, and the second-person standpoint, and then arguing for a similar conceptual tie between responsibility or accountability, and moral obligation. These second-personal conditions, he argues, are not reducible to other, non-second-personal conditions. "Propositions formulated only with normative and evaluative concepts that are not already implicitly second-personal cannot adequately ground propositions formulated with concepts within the circle [of second-personal concepts]" (Darwall 2006, p. 12).

After establishing the essential second-personal nature of moral obligation he undertakes to show that moral demands, and second-personal claims generally, are justified, in the sense that their purported normative authority is real. In other words, Darwall argues that rational agents in general[18] are accountable to each other as free and equal rational beings (Darwall 2006, p. 300). Each is in the same legitimate position to make demands on every other, according to the arguments in the latter part of Darwall's project.

B. Obligation as accountability

Darwall's own arguments for the claims, which follow below, of the irreducibility of second-personal reasons and the second-personal nature of obligation are a good deal more complex and detailed than I am able to replicate. However, I believe that the main ideas and overarching arguments can be grasped sufficiently from my short presentation so that one can see the power of this interpretation of moral obligation. I hope to at least be able to motivate my fusion of his view with my own view of practical normativity on the grounds that the result is a sophisticated theory of moral obligation with many merits and virtues.

Darwall's story begins with an example. Say that some guy steps on your foot. This causes you pain, and the question arises what kinds of reasons the man has to remove his foot from on top of yours. He might have sympathetic reasons. That is, he might have reasons that stem from his own desire for you to be free of undue pain. He might also have an *agent-neutral* or *state-of-the-world-regarding* reason to remove his foot, which just exists as a reason to make the world better. This latter type of reason

[18] It is not clear that Darwall means to show something about *every* rational agent, regardless of the agent's empirical dispositions. He may be assuming something to be generally empirically true of human beings, which helps establish the appropriate relations. This is a point I will return to at the end of the chapter.

"would not be essentially for him as the agent causing another person pain. It would exist, most fundamentally, for anyone who is in a position to effect your relief and therefore for him, since he is well placed to do so" (Darwall 2006, p. 6). Concerning this type of reason, the man would only have it because he can so easily remove his foot, whereas, other things equal, some bystander would have a harder time removing his foot without creating other problems (like doing the man some harm, or potentially angering a crazy stranger). You demanding that the man succumb to such a reason would amount to your trying to point out that a reason exists for him to remove his foot, and would not depend on your having any competence or authority to *give him* a reason to remove his foot.

A different kind of reason would be that,

> you might lay a claim or address a purportedly valid demand. You might say something that asserts or implies your authority to claim or demand that he move his foot and that simultaneously expresses this demand. You might demand this as the person whose foot he is stepping on, or as a member of the moral community, whose members understand themselves as demanding that people not step on one another's feet, or as both (Darwall 2006, p. 7).

This is a *second-personal reason*. Second-personal reasons are fundamentally agent-relative and not agent-neutral. They exist for particular persons because of who they are and what relationships they are in with the addressers of second-personal claims. For the case at hand, it is clear that a second-personal reason for the man to remove his foot would be the most significant reason in the situation. Even if there exists an agent-neutral reason for him to remove his foot, this would not ground your right to demand

that he comply with the reason and to "enforce" the demand with whatever appropriate means are at your disposal (which might just be blame and resentment). Reasons, whether agent-relative or agent-neutral, cannot ground the authority to demand compliance, unless they are second-personal in form. An explanation of second-personal authority is the only way to establish accountability and legitimate compliance demands.

Darwall has it that there is an interdefinable circle of second-personal concepts which cannot be reduced to concepts outside of this circle. The key elements of this circle are second-personal authority, valid claim or demand, second-personal reason, and responsibility to (Darwall 2006, p. 12). Having second personal authority implies having the standing to make a valid claim or demand on someone. Making a valid claim or demand creates a special kind of reason for the addressee to comply, and having a second-personal reason to do something means that one is responsible for complying. Going the other way, being responsible or accountable implies that someone has the authority to hold one accountable. These are all conceptual relations between second-personal concepts.

If you validly demand that I remove my foot from on top of yours, you are making a direct, second-personal claim against me. This is not merely advice you are giving me to the effect of making me aware that some third-personal reason exists for me to remove my foot. The reason you give me for removing my foot depends on your relation to me and not on my relation to various "facts and evidence as they are anyway" (Darwall 2006, p. 12). Your authority in giving me a reason is second-personal. It stems from our relation to each other and your position to make a demand of me that I am accountable to you for complying with. All second-personal reasons are also first-

personal – they are agent-relative, and so stem from the first-person standpoint – but not all first-personal reasons are second-personal. Many first-personal reasons lack the aspect of addressing someone, which brings in the second-person standpoint (Darwall 2006, pp. 9-10). As Darwall points out, this does not mean that second-personal reasons do not need justification or explanation (Darwall 2006, p. 13). It is simply a claim about what they are and what they are not.

Having set up the building blocks of the second-personal concepts, the next step for Darwall's thesis of moral obligation as accountability comes in two stages: (1) An argument that accountability is second-personal. (2) An argument that moral obligation springs from accountability. The conclusion can then be drawn that moral obligation is second-personal in the sense that it arises from second-personal relations. Second-personal relations are essential to moral obligation, and moral obligation is essentially second-personal.

To get the first conclusion, that accountability is second-personal, Darwall starts by looking at P.F Strawson's concept of "reactive attitudes" which "respond to 'transactions' between individuals" (quoting Darwall 2006, p. 66; Strawson 1974, 1-25). Reactive attitudes can be *participant* or *impersonal*, which correspond to direct responses by individuals in the transactions, or directly on their behalf, and responses from a disinterested perspective, as from the moral community. Participant reactive attitudes include gratitude, resentment, forgiveness and hurt feelings. Impersonal reactive attitudes include moral indignation, disapprobation, and guilt (Darwall 2006, p. 67).

Both Darwall and R. Jay Wallace (Wallace 1994) argue that responsibility should be understood in terms of warranted reactive attitudes. Wallace's argument takes up the

idea that "holding people responsible is a form of deep assessment that goes beyond mere description of what an agent has done, insofar as it involves a liability to the responses of blame and moral sanction" (Wallace 1994, p. 82). That which "goes beyond mere description," Wallace holds, can only be properly captured by warranted reactive attitudes (or "reactive emotions" as Wallace calls them). Responsibility is being blameworthy for non-compliance, and judgments of blameworthiness are equivalent to judgments of the appropriateness of certain reactive attitudes.

Darwall's argument is related, but it spins off of what he calls *Strawson's Point*: "Desirability is a reason of the wrong kind to warrant the attitudes and actions in which holding someone responsible consists in their own terms" (Darwall 2006, p. 15). This is to say that state-of-the-world-regarding reasons are reasons of the wrong kind to establish responsibility. A consequentialist may claim that desirable outcomes justify holding people responsible for their undesirable actions (in the face of a determinism-based objection), but Strawson argues that this is a "reason of the wrong kind" (to use Darwall's phrase) or "not . . . even the right sort of basis" (in Stawson's words) (Strawson 1974, p. 4). Whether or not consequentialism can justify responding to undesirable behavior in harsh ways, it cannot justify holding people responsible in a pure sense.

Darwall's final move for this stage in the argument (that accountability is second-personal) is to argue that the reactive attitudes are forms of second-personal address. This argument has the simple form of pointing out how each of the examples of reactive attitudes involve second-personal address. I will only look at a couple, since I do not consider this to be a very interesting or contentious part of the overall argument. Take guilt, for example. To feel guilty is to feel that blame is appropriate, and that there is

some authority in someone to level blame at oneself. Thus, feeling guilt implies feeling "as if one has the requisite capacity and standing to be addressed as responsible" (Darwall 2006, p. 71). Looking at the impersonal reactive attitudes, moral indignation and disapprobation, we see that these are part of what it is to hold someone accountable – to feel that they are rightly held responsible. These attitudes are what we feel when we "feel we can demand, as members of the moral community, that people act in certain ways" (Darwall 2006, p. 74). Hence, the reactive attitudes, with these as representatives, are forms of second-personal address. It follows from the above arguments that responsibility, and with it, accountability are second-personal – they arise from, and exist in the second-person standpoint.

The second stage in Darwall's project of arguing that morality is second-personal is to argue that moral obligation springs from accountability. Down the line, this allows Darwall to make out his conception of morality "as equal accountability." Darwall sees this stage of the argument as being crucial to the overall project of analyzing morality's normativity, since he takes the moral obligation as accountability argument to show that an essential aspect of morality's bindingness is that it is tied to accountability. Indeed, Darwall claims,

> Even if it were possible, consequently, to account otherwise either for moral obligations' invariably purporting to provide superior normative reasons or their actually doing so, it would still be impossible thereby to explicate the distinctive hold or bindingness that moral obligations purport to have. This can only be done in second-personal terms (Darwall 2006, p. 100).

I thoroughly analyze later on, the extent to which this, and related claims, are true. In the sense that second-personal reasons have to be explained in a way that obligations "go through" them, I accept this. But, in the sense that the accountability is itself the original source of normativity, I do not.

Darwall's argument for his conception of obligation as accountability begins by arguing that there is strong precedent for the conception in the writings of prominent philosophers (Darwall 2007a, pp. 123-125). He specifically cites Strawson, whose analysis of moral responsibility is Darwall's starting point for his account of the second-personal nature of morality. Strawson, for example, includes "a sense of obligation" as one of the reactive attitudes (Strawson 1974, p. 15), and he lumps the fate of moral obligation in with the fate of responsibility in the debate over determinism (Strawson 1974, pp. 2-3). Darwall also cites Mill as clearly holding this view and quotes him saying as much: "We do not call anything wrong unless we mean to imply that a person ought to be punished in some way or other for doing it" and "It is a part of the notion of Duty in every one if its forms, that a person may rightfully be compelled to fulfil it" (Mill 1861,1979, p. 47).

Thinking about these authors' views may be convincing in itself. However, Darwall has another argument, which seems to me the more profound one.

> Nothing depends, of course, on whether we use the words 'wrong' and 'moral obligation' in the way Mill and [others] say we do. We could use these words more broadly to include moral ideals or goals. However, if we did, we would still need terms to refer to the idea to which these thinkers point, namely, the part of morality that concerns that for which we appropriately hold one another

> responsible. . . With this understanding, therefore, I shall henceforth use 'wrong' and 'moral obligation' in a Millian way as implying accountability-seeking demands (Darwall 2007a, p. 125).

Darwall's argument here is a form of what I have elsewhere called "pragmatic explication" (Hammer 2008, 183-199). This argumentative technique looks to what role a term is needed to play in various philosophical enterprises and analyzes what meaning is needed for the term in order to play the role. For example, to greatly oversimplify, I argue in the same place that 'explanation' is needed to play the role of denoting the endeavor to provide *intellectual control* over a scenario or phenomenon to be explained, where this intellectual control is understood as hypothetical practical control, which can be reduced to a set of specific conditions like recognizability, manipulability, and non-replaceability. By pragmatic explication, this means that we can analyze 'explanation' in terms of the search for intellectual control, and thereby provide an important and relevant definition of 'explanation' that is applicable for any use of the term 'explanation' drawing on this role. Thus, we can use such a definition to show the common sense in which the natural sciences, the social sciences and history (which can be considered as its own category) all explain why something happened.

In the present context, we can see the need for a term that captures the accountability aspect of morality. This accountability aspect, according to which a moral agent can be accountable to others in a sense that gives others the authority to exact compliance with certain demands, would clearly be at least a *kind* of moral obligation. But we can also see it as more than that, because 'moral obligation', in the sense of that which one can appropriately be held responsible for, plays the role of denoting something

that is fundamental to a minimal sense of morality. The modern idea of morality includes that which people are responsible for.

Although he does not present it this way explicitly, we can see Darwall's follow-up argument to his need-for-a-term argument as support for the relevant pragmatic explication of 'moral obligation' (Darwall 2007a, pp. 126-128). Darwall turns to two debates in the literature on moral theory to show that the obligation as accountability understanding can help clear them up and is presupposed by (at least some of) the participants in the debate. These two points, I presently argue, can be applied to our pragmatic explication of 'moral obligation' to show the indispensable work the obligation as accountability interpretation does in moral theory. The first point concerns the debate over whether act-consequentialism is too demanding (Darwall 2007a, pp. 126-127). It is argued by some that act-consequentialism is too demanding, and should therefore be dismissed or considered inadequate to capture right and wrong. Darwall interprets the issue as whether act-consequentialism recommends more than can be authoritatively, or second-personally demanded by the moral community – whether maximizing overall net good is doing more than one is accountable for. Now, this interpretation may not settle the issue of whether act-consequentialism is too demanding, but it does do a great job of clarifying what is at issue. In the absence of the accountability interpretation, there is no clear standard of what "too demanding" would mean.

The second debate Darwall points our attention to "concerns morality's purported authority, that is, whether moral obligations are categorical imperatives in Foot's sense of always necessarily giving (conclusive) reasons for acting" (Darwall 2007a, p. 127). We want to say that a moral demand always purports to be authoritative in the sense that one

merits blame for not meeting the demand when one is subject to it. This can be understood through the accountability interpretation, since accountability relations carry with them the idea of blameworthiness for non-compliance. When Philippa Foot looks for a sense in which moral demands differ from the demands of etiquette, since both seem to always apply (Foot 1972, 305-316), Darwall points out that the difference can easily be captured by accountability (Darwall 2007a, p. 128). We are accountable for moral demands, but not the demands of etiquette, unless there is some special relation that gives etiquette a moral quality. I later argue that accountability and second-personal reasons cannot explain their own normative authority, but the present point is that, conceptually, they provide a way of understanding the unique kind of authority that morality purports to enjoy.

So, moral obligation should be understood in terms of accountability, because that is required by the work that the term 'moral obligation' needs to perform. What people are responsible for is a conceptual component of morality. This component is second-personal, as was argued above. Since the role that 'moral obligation' is needed to play in moral theory overlaps with what people are responsible for, the best way to understand moral obligation is in terms of responsibility or accountability. Therefore, we can draw the conclusion that moral obligation, pragmatically understood, springs from second-personal accountability.[19]

C. Darwall's constructive arguments

[19] Strictly speaking, I should be able to use pragmatic explication for all explicative arguments, since, it seems to me, this is the only foolproof method of informative explication. This would mean that all of my arguments in Part II should, in principle, be capable of a pragmatic-explication formulation. I believe that such a formulation is available for all of those arguments, but it is not necessary, since I analyze concepts that have very fixed roles. Agency is just initiation of action through a unified rational process, for example. Since it would be unnecessary to use pragmatic explication for my Part II arguments, there is no point in increasing their complexity through such formulations. It is always more complicated to provide an analysis through pragmatic explication.

The second major component of Darwall's project is to vindicate morality's purported normative authority. This amounts to an attempt to explain how it is that anyone (and everyone) actually has moral obligations and is accountable to one another. Under Darwall's scheme, this has three components:

(1) A transcendental argument for the presuppositions of second-personal address. This involves arguing that to use second-personal address at all, various conditions must be met which jointly amount to something like imposing morality on oneself.

(2) Vindication of the second-person standpoint. Whatever may be required for second-personal address to occur, the question stands as to whether anyone ever succeeds at performing second-personal address and whether doing so is in any sense unavoidable. Darwall needs to argue that the second-person standpoint has some kind of legitimacy.

(3) Establishment of morality's authority or normative supremacy. This does not necessarily involve demonstrating and explaining normative supremacy, but it does require an assessment of what is needed here and what role this plays in morality's normative force. Morality must have some normative force, even if it is not always normatively supremely authoritative.

The first component of Darwall's apparatus for the reality of obligation is the alleged presuppositions of the second-person standpoint. Darwall argues that there is a complex set of conditions that must be presupposed in order for anyone to address

another second-personally. In order for me to address you in a purely second-personal way, and not to merely coerce you, Darwall has it that I must presuppose a list of things about you, me and our transaction. If I do not presuppose these things, then I am not really addressing you second-personally. I do not need to be consciously aware of the presuppositions – that would rule out successful address of this kind immediately, because no one thinks of all the conditions Darwall cites when they try to address one another. The presuppositions are not a matter of choice, but of logical form. Just as I cannot go to the doctor (*de dicto*) without presupposing that I am headed toward a place where there is a person of various qualifications, so also I cannot perform second-personal address without presupposing, according to Darwall, that I and the addressee are free and equal in a second-personal sense.

The part of the argument that I want to focus on, because it is the part I criticize, is Darwall's move to the equal authority of addresser and addressee. Darwall argues that in order to address second-personally I must presuppose that my addressees "can freely and rationally determine themselves by the addressed reasons" (Darwall 2006, p. 269). Darwall argues that a consequence of this is that "addressing a demand always presupposes a distinction between legitimate . . . ways of relating to someone directively . . . and illegitimately coercing him" (Darwall 2006, p. 271). This presupposed distinction, Darwall believes, implies presupposing that coercion of the addressee would be illegitimate and would violate the addressee's authority as a free and rational person. Darwall concludes, "So any address of second-personal reasons must presuppose that addresser and addressee share a (second-personal) authority to make demands of one another as free and rational persons" (Darwall 2006, p. 274). This is a very important

argument for Darwall. This is how he supposedly establishes the equality of persons as a presupposition of any second-personal address. Of course, I have serious objections to this argument, which are made out later on.

After having argued for mutual accountability as a presupposition of second-personal address, Darwall needs to establish some kind of legitimacy for the second-person standpoint all together. Whether or not anyone can and does, or must take up the second-person standpoint and the presuppositions it entails, and whether or not there exist second-personal reasons which bind moral agents are questions which remain open at this stage.

Darwall has a two-part strategy for addressing this need. He first argues that it appears psychologically impossible or extremely difficult to avoid taking up the second-person standpoint (Darwall 2006, pp. 277-278). This is supposed to show that any sense in which the second-person standpoint is optional is really irrelevant, since everyone we are worried about must, in fact, take it up. While this point, as it stands, seems a little funny, I do not criticize Darwall for this, since ultimately I take moral obligation to be empirically contingent anyway. Furthermore, it seems to me that Darwall is very close to pursuing a constitutivist line himself. If he argued that the second-person standpoint were constitutive of moral agency (which he seems close to arguing) and that the second-person standpoint, therefore, governs moral deliberation (or something like this), he would have a kind of a constitutive view. In such a case, Darwall would not need to worry about the empirically contingent aspect of the second-person standpoint. However, he would still, I think, have the problems with normative supremacy I discuss

in detail later on, because he fails to show that the second-person standpoint has some special place within the overall deliberative standpoint.

To address the issue of taking second-personal reasons seriously, Darwall has a different kind of argument, which I only outline. This argument takes a kind of practical necessity line, which, as a general argument strategy for normative reasons, is more fully made out by David Enoch (Enoch 2007). Darwall gives a version of this argument both in *The Second-Person Standpoint* and in his "Reply to Korsgaard, Wallace, and Watson" (Darwall 2007b, 52-69). The argument runs that deliberation requires practical reasons, and we have as good reason to accept second-personal reasons as we do any kind of practical reasons on all our major metaethical theories. Therefore, "we should conclude that, for practical purposes, at least, second-personal reasons grounded in our common dignity really do exist" (quoting Darwall 2006, p. 299; Darwall 2007b, p. 60). This ties into the particular points about the presuppositions of second-personal address by the way that second-personal reasons are supposed to be legitimated by the "major metaethical theories." I do not criticize this argument in any detail, but I will point out that it will not help him solve the various theoretical problems for his system that lay out, because it is simply a practical argument which defends against metaphysical skepticism about second-personal reasons. My objections do not rely on this kind of skepticism.

This, at last, brings us to the most important issue for Darwall's project: whether and to what extent he can claim normative supremacy or authority for second-personal reasons and accountability. Darwall clearly wants to demonstrate the normative supremacy thesis for obligation as accountability. First of all, he of course claims that moral obligations "purport" to "give agents conclusive reasons for acting" (Darwall 2006,

p. 26). He supports this claim at length, at least throughout Chapters 4 and 5. This line of reasoning is easy to follow. If I hold you accountable, and say that you should not have done X, I assume that you had conclusive reason not to do X. If you show some reason to do X, that speaks against your responsibility not to have done it. Say that I expect you to pick up some tickets for us. If you don't get the tickets, I will probably hold you responsible by blaming you and feeling a little resentment toward you. If, however, you have good reason, an overriding reason not to have picked up the tickets, then my blame and resentment would be inappropriate. You would not be responsible (at least not fully responsible) for not getting the tickets. You are only responsible and worthy of blame if you had conclusive reason to get the tickets.

To put moral obligations on the straight and narrow, he needs to show that they do in fact give conclusive reasons for acting. The Supremacy Thesis, or an approximation of it (e.g. "Moral demands are generally supremely authoritative"), is necessary to establish the relevance of all of Darwall's conceptual, second-personal-essence arguments. It is also necessary to vindicate the whole practice of holding people accountable and responsible, as Darwall himself argues. If there is no basis for the idea that second-personal demands of some kind give conclusive reasons for acting, then the whole practice of morality and of holding people responsible is jeopardized. There may be nothing sensible in the practice under such circumstances.

Darwall also explicitly says that he argues for this:

> In Chapters 10 and 11, I take on [the question of whether the Supremacy Thesis holds] directly and argue that a conception of morality as equal accountability can be grounded in a theory of practical reason in a way that vindicates the claim that

> moral agents always necessarily have conclusive reasons not to violate moral obligations (Darwall 2006, p. 97).

But does Darwall really make such an argument? The closest he comes appears in a footnote.

> Have I shown that such reasons are supremely authoritative? If the argument of Chapter 10 is sound, we cannot avoid presupposing the equal dignity of free and rational persons when we address second-personal reason of any kind. Since attempting to hold one another accountable through second-personal address itself apparently involves an assumption of the supremacy of the accountability-seeking second-personal reason (see Chapter 5), it is hard to see how to avoid presupposing the supremacy of reasons of equal dignity (Darwall 2006, p. 300, n. 1, reference to the author's chapter).

The reasoning here draws on the claim that accountability, and with it obligation, conceptually presupposes conclusive reasons. Darwall thinks that complex presuppositions amount to a basis for the supremacy of accountability-seeking reasons.

I see this argument as a bit of a flop. Even if "it is hard to see how to avoid" were the right kind of argument to make, "presupposing" seems to be out of place here. What we are looking for at this point is an argument that actually shows that moral obligations have some kind of normative supremacy, not another argument that conclusive reasons are assumed when we make use of accountability concepts. To shorten what Darwall says: since using accountability concepts "involves an assumption of . . . supremacy," we must presuppose supremacy (skipping the "hard to see how to avoid" business). I surely grant that presupposing supremacy follows from assuming supremacy, but I do not see

how an argument like this is supposed to help. It seems that what Darwall is trying to do is make a slide on the *mutual* aspect of accountability relations. Consider this line of reasoning: "If I demand that you not step on my foot and hold you accountable for not doing so, then I assume that you have conclusive reason not to step on my foot. Similarly, you, in holding me accountable for not stepping on your foot, assume that I have conclusive reason not to step on your foot. So, we each assume the other has conclusive reason to abide by second-personal reasons. So, we all agree that second-personal reasons are supremely authoritative." This line of reasoning is obviously faulty. Jumping from a mutual assumption of a particular point of conclusive reasons *for the other person* to the conclusion that we all have conclusive second-personal reasons is entirely unwarranted. I do not see a better way of reconstructing the argument. It simply does not work.

IV. The shortcomings of this view

A. Problems with Darwall's constructive arguments

Darwall's constructive arguments – that is, his arguments which are supposed to substantively show the reality of moral obligation – while creative, subtle and insightful, are faulty and unsound. To show this, I first make use of Gary Watson's criticisms of Darwall, which do a very good job of cutting right at a weak move that Darwall makes which is essential to his project of establishing a foundation for moral obligation and contractualism (Watson 2007, 37-51). The part of Watson's critique I wish to draw on is rather straightforward once it is laid out clearly and should do a good job of showing that Darwall's view needs some outside help.

While Darwall tries to go down a very purely second-personal road in order to make his constructive arguments, a path which is encouraged, I think, by his sincere dedication to the second-personal nature of morality, this road does not bring him to his destination. He needs to take a different route. In the next subsection, after having presented Watson's critique and an extension of it, I argue more generally that Darwall's problems cannot be fixed by any means implicit in other things that he says. In order to solve his problems, Darwall needs help from outside of the purely second-personal approach he is determined to take. He needs to go back to a general, perhaps properly first-personal, account of normativity (like mine, happily enough!), but I get to that issue, the remedy, later in the chapter.

Watson's objection focuses on the presuppositions of the second-person standpoint that Darwall wants to establish. Specifically, Watson looks at the idea that using second-personal address commits one to viewing others as having equal second-personal authority to oneself. Recall that Darwall argues from the need to presuppose an addressee's ability to accept one's addressed claim freely and rationally, to the need to presuppose that one is not "illegitimately coercing" the addressee, but addressing him directively as free and rational. Darwall takes this to imply that one commits oneself to standards of accountability to the addressee by subjecting oneself to norms of warranted demands that make one's address legitimate and directive rather than illegitimate coercion.

The overarching problem with this argument is that the requirement that the addressee be able to rationally accept a demand (as legitimate) does not imply that the addressee has some kind of authority which is the same as the addresser's authority to

make the demand. If my demand of you presupposes that you can freely and rationally accept the demand, that does not mean you have to be able to impose the demand on yourself by the selfsame authority I enjoy in making the demand of you. There does not need to be any common authority in this sense.

While this is the overarching structural problem with the argument, we can also see specifically where the gap is in Darwall's official argument. Presupposing that I am not coercing you is not the same as presupposing that you have a right that I not coerce you or treat you in some way that I cannot justify from a shared perspective. Watson focuses on this problem, although the way he phrases it is a little less pointed than it could be. Watson questions whether "being improperly subjected to certain demands entail[s] being *wronged* by being subjected to those demands?" (Watson 2007, p. 45). Watson goes on to say that the inference would rely on a substantive normative principle "that we owe it to second-personally competent agents to treat them only in ways that we can justify to them" (Watson 2007, p. 45). Further, this principle does not seem to be presupposed by second-personal address, and, in any case, Darwall has no argument for it. This is a way of making out the objection, but it seems to give Darwall a little more than he has earned. There is no need to grant that one commits oneself to standards that could make one's demand "improper." I might put a demand to you to give me five dollars. If I am not putting this to you in a way that presupposes you can rationally accept it, then I am not second-personally addressing you; I am bullying you for five dollars. If I am, then I am not subjecting you to it improperly. In order for you to actually be able to accept my demand for five dollars rationally, the circumstances might need to be right – like you owing me five dollars already. But the issue here is what I

presuppose in my demand. I may presuppose that you can accept the demand of your own choosing in some sense, but this does not imply that I impose standards of accountability to you on myself.

We might add to Watson's objection another point, which Watson does not seem to make, although Watson does say a lot more than what I have canvassed above. Not only is there a problem with Darwall's equal mutual accountability between addresser and addressee, there is also the problem that there is no established sense in which the addressee is representative of everyone. Even if I commit myself to accepting your equality when I address you, this does nothing to commit me to the equality of anyone else. For all Darwall has to offer by these presupposition arguments, I could address a handful of people but then treat others with complete disrespect and callousness. I do not see an argument for a general system of moral obligation that includes duties to everyone.

There are two ways in which Darwall might have a reply. He might say that we cannot help but address "the whole community" and commit ourselves generally in that way, or he might try to find help in contractualism. Darwall seems to be relying on the first idea, that we cannot help but address the whole community. Darwall clearly relies on his conclusion that we are committed to the equal authority or "dignity" of all in his formulation of a foundation for contractualism that he explores in Chapter 12 (Darwall 2006, pp. 300-320). Furthermore, if the conclusion of general equality did not come first, then he would not be supplying a foundation for contractualism at all, but would rather be looking to contractualism for a foundation for his conclusion.

So, Darwall must be relying on the idea that we cannot help but to address all, and to commit ourselves generally to the equal authority and dignity of all. But even if we

grant Darwall the claim that it is nearly impossible to avoid blaming and resenting and holding people accountable, and so to take up the second-person standpoint, that does not seem to help in getting the generality he needs. I will be the last to complain that moral obligations, according to him, are only contingent, but the problem goes beyond that. The problem is that, particularly in the case of those who have little or no power over me, there seems to be no pressure on me to take up second-personal attitudes toward them and to assume their equal dignity. In order for the question of me holding Harry accountable to arise, Harry must have some kind of power over me. Otherwise there is just nothing for me to demand of Harry, and so no reason for me to take up second-personal attitudes toward him. Although the moral community may demand of me that I still treat Harry with respect, this only puts us in a gray area where Darwall does not seem to have any tools. Perhaps contractualism can help us out with this to the extent that it offers an answer to what I owe the moral community and whether or not it can demand my respect of Harry. But that does not mean that I am committed to the equal authority of all or that contractualism can be founded on the idea that I have such a commitment. While I believe a lot more could be said about this, going further into this issue would take us too far away from my central line of argument. Since I do need to press this objection further, I leave it here and move on to what I take to be the primary problem for Darwall's approach – the question of the normative supremacy of accountability.

B. General problem with normative supremacy and deliberation

In the previous subsection, I argued that Darwall's constructive arguments do not work. He does not succeed in showing "that the demands we have standing to make of one another as persons generate valid normative reasons for acting" (Darwall 2006, p.

300). I now want to argue that there is no immediate cure for his ills. Darwall cannot provide a normative foundation for morality as accountability without going outside of the purely second-personal approach. In order to ground accountability, we need to look beyond the second-person standpoint and consider a more general scheme of practical normativity. Of course, just such a general scheme of practical normativity is what I provided in Part II. Later in the chapter, I show how my theory of practical normativity can be fused with an account of moral obligation as second-personal accountability. In the next section, I present a puzzle which lays out clearly what the difficulty is in fusing the two accounts, and which puts this tension in terms of a broader conflict in moral theory between first-personal accounts of practical rationality and second-personal accounts of social morality. Thereafter, I show how this conflict can be overcome and my theory of practical normativity can be combined with a second-personal account of obligation, which does justice to Darwall's purely conceptual arguments I discussed in III.B.

At this point, I want to show what is wrong with trying to ground moral obligation from a purely second-personal perspective, without using a general scheme of practical normativity. Let us, for a moment, forgive all of the shortcomings in Darwall's constructive arguments, and allow that he has shown the legitimacy of equal common second-personal authority among all members of the moral community – that is, everyone who can be addressed second-personally or who can address second-personally. As I pointed out at the end of Section III, what appears to be Darwall's final argument for the Supremacy Thesis does not work, although there is clearly a need for such an argument. Even if Darwall cannot show that moral obligations are always supremely

authoritative (something many moral theorists contest), he needs to be able to give some kind of conditions for when moral obligations would provide conclusive reasons or how they fit into a general scheme of practical normativity.[20] Without this, Darwall does not really have an explanation of the normative force of accountability at all. We need an explanation of the role that second-personal reasons play within practical normativity. Just showing that we presuppose things in a certain way, does not imply that there is genuine normative force behind the presuppositions. If I must presuppose that you have the authority to give me reasons, that does nothing to determine how these apparent reasons are supposed to fit into my deliberation.

Following up on this line of thinking, the problem with a purely second-personal approach can be captured by the following simple argument.

Argument 8

(1) Normative supremacy must be general deliberative supremacy – normativity exists, fundamentally, in the deliberative sphere.

(2) Supremacy must range over deliberation in general, not just a particular area of it – if it does not range over the whole of the deliberative field, it is not really supremacy.

(3) Deliberation in general is first-personal, but not second-personal (in general) – as Darwall notes, the second-person standpoint is also first-personal, but the converse is not true (Darwall 2007b, p. 55).

[20] Arguing that second-personal reasons are just as acceptable as other reasons on a constructivist view, or that our intutions for second-personal reasons are just as strong as our intuitions for other kinds of reasons (acceptability on a recognitional view), as Darwall does Stephen Darwall, *The Second-Person Standpoint: Morality, Respect and Accountability* (Cambridge: Harvard University Press, 2006), , pp. 293-299., does not do the job.

(4) Hence, normative supremacy must come from the first-person standpoint, but not the second-person standpoint.

The first premise points to the centrality of deliberation for normativity. Now, someone might reject everything I argued in Chapter 1 about the need for explanation and the ineptitude of classical realism to provide or even fit with an explanation. One might take up a version of classical realism, or a "recognitional theory" of practical reason as Darwall calls it, following Cullity and Gaut (Cullity and Gaut 1997, p. 4), according to which "there exist independent normative reason facts, and whether a person or deliberative procedure is rational depends upon whether she or it adequately registers or recognizes these" (Darwall 2006, p. 292). But, even in this case, while normative reasons would in some sense exist outside of the deliberative standpoint and standpoints parasitic on it (like a third-person evaluative standpoint which tracks normativity according to reasons from the deliberative standpoint of the acting agents[21]), reasons would only have meaning in the context of such standpoints – that is, they would only have meaning as they played into deliberation. Ultimately, I am not concerned with how my arguments play out on a recognitional theory since I reject such approaches to normativity. However, it is worth pointing out that this problem remains for Darwall even were he to embrace such a theory.

The second premise I take to be fairly straightforward. Of course, if we grant that normative supremacy must be deliberative supremacy, then supremacy must range over the whole of deliberation – over the deliberative standpoint as such. The whole of the

[21] For more on this relationship, see my discussion of the Assessability Thesis and of constitutivism in Chapter 2.

deliberative standpoint makes use of reasons. So, the supremacy of reasons claim must be a claim that some specified set of reasons have authority over all other reasons which means over the whole deliberative standpoint. This really just follows from the first premise and the meaning of the supremacy claim.

The third premise would seem to be controversial only according to more radical views of the social nature of morality and normativity. If we thought that to deliberate at all was to do something social, then we would have to find fault with the third premise. But, even if, for example, we accepted the view that deliberation requires the use of a complex linguistic capacity which is itself inherently social, it would still not follow, in any relevant sense, that deliberation itself is social or second-personal. Deliberation would still fail to be second-personal in Darwall's sense that it "depends on presupposed authority and accountability relations between persons" (Darwall 2006, p. 8). Indeed, if premise (1) were true, and deliberation were second-personal in Darwall's sense, then the Supremacy Thesis would be entirely trivial and the whole point of moral obligation as accountability would be lost.

Therefore, normative supremacy must come from, and occur in the first-person standpoint as such, which is not the second-person standpoint. We need a scheme of practical normativity in general, in which we can embed the notions of accountability and second-personal reasons. Whether or not we can show that accountability really is normatively supreme in all cases, we need to be able to place accountability in a larger picture of how normativity operates within the first-person standpoint. Doing this is how we can validate morality's purported normativity to whatever extent it should be

validated, and how we can establish the role accountability plays in practical normativity as such.

Thus far, I have argued that Darwall's notion of moral obligation as accountability is correct, but his constructive scheme for establishing morality's legitimate normative role and its normative supremacy is a failure. I said at the beginning of this chapter that I would be using Darwall's ideas on the second-person standpoint to help construct a theory of morality based on my theory of general practical normativity I developed in Part II. I am now in a position wherein I must explain how morality as accountability connects with practical normativity as systematization in order to make sense of my arguments to this point and to put forward a theory of morality. If moral obligation must be understood in terms of accountability relations and their normative hold on us, then either I must vindicate the normativity of accountability relations by fusing Darwall's interpretation of obligation with my account of practical normativity, or I am left with a moral error theory – something I do not by any means endorse. In the next section I look at the difficulty of fusing my theory with Darwall's and the general tension between the second-personal nature of morality and the first-personal nature of normativity. In the section thereafter I carry out the promised fusion and explain how the tension between the first- and second-person standpoints, which are constitutive of morality, can be resolved.

V. The puzzle

A. The puzzle for my view – how to meet Darwall's conditions

In this section I lay out the difficulty of reconciling the second-personal nature of morality with the first-personal nature of practical normativity. I do this by first stating the problem for my own view: how can Darwall's second-personal conditions be met on my theory of normativity? I explain more thoroughly what the difficulty inherent in this project is. Thereafter, I state a more generalized version of the puzzle and what the conditions are for this puzzle to arise. In the next section, I present my solution to the puzzle and resolution to the tension between the first-personal and second-personal aspects of morality.

According to my theory of practical normativity, all rational agents have the same standard for right action – systematicity. How constructed practical principles figure in to one's systematic scheme of principles is what determines whether or not they should be acted on. Some principles have authority over other principles because they enjoy a more privileged position in the systematic scheme. Of course, no one is ever consciously aware of the entire scheme of principles they operate on, but this scheme lives in the dispositions and past choices of the individual agent. Whenever someone makes a new (rational) choice, that agent will *de facto* use some pattern of reasoning which draws on the systematization he or she accepts and will thereby incorporate the new principle into the scheme. Say I wear a helmet while riding my bike *because* I take the dangers of the road more seriously than I used to and I am no longer as concerned with the inconvenience and "emotional discomfort" wearing a helmet can bring. Then this means I incorporate a principle like, "I wear a helmet when I ride my bike through the city" into my scheme. It will occupy a position whereby it is a consequence of sorts of subsuming principles concerning my attitudes toward safety. Further, I modify principles

concerning my attitudes about the inconvenience and discomfort of wearing a helmet, to resolve the conflict that forced me to debate with myself. This, roughly, is how systematization works.

Now, this aim at, and process of systematization is entirely first-personal and is a general scheme for how the whole first-personal practical standpoint works. It provides standards for what deliberation is, and what it tries to accomplish.[22] So, it certainly has the potential, and is in fact geared to address questions of normative supremacy. However, it remains a first-personal and not a second-personal process. As we saw, moral obligation must be second-personal in the sense that obligations must arise through the bindingness of accountability relations. Furthermore, these accountability relations are not a form of, and cannot be reduced to strictly first-personal or non-second-personal relations or conditions.

B. Statement of a generalized version of the puzzle and its preconditions

We are now in a position to present a more generalized version of the problem, which can be stated as a puzzle for moral theorists who endorse three simple ideas. These ideas are (1) that commonsense moral obligations are real, at least roughly speaking, (2) the normative supremacy of obligation, or at least an authoritative position for obligation within the overall scheme of practical normativity, and (3) an explanatory approach to moral theory. These three ideas are by no means universally accepted, even among those who would claim to take morality seriously. However, they are fairly common and that makes this puzzle relevant for moral philosophy in general.

[22] Of course, there could be other kinds of mental processes which led to behavioral outcomes, and which in turn we might call "deliberation." But, whatever that has to do with the first- and second-person standpoints, it would not be a form of rational, or practical deliberation as I discussed in Chapter 2.

Furthermore, these conditions are appealing in the sense that they represent some of the key components in taking morality seriously and trying to incorporate it into a developed philosophical framework. The first idea is just that of buying into morality in some general way. If we accept that we really do have obligations and that these obligations are loosely what we think they are – maintaining respect, kindness and some measure of generosity – then we accept condition (1). The second idea may be more controversial, and obviously presupposes that we do have obligations, but as I have argued, using Darwall's ideas, moral obligations purport to be supremely authoritative. They arise from our holding one another accountable, and holding someone accountable implies an assumption of conclusive reasons. So, accepting condition (2) is also part of taking morality seriously. Condition (3) arises, not so much from taking morality seriously, but rather from taking *theoretical* rationality seriously. If we don't look for explanatory unity in our worldview, then we are not really being serious about constructing a philosophically sound outlook. The three conditions which lead us to the puzzle may have their setbacks, but they do form part of what it would be to take morality seriously and try to fit it into a developed philosophical framework.

The puzzle

Argument 9

How to reconcile the following three true propositions.

a) Moral obligations must arise *through* second-personal relations. (Section III.B)

b) The fundamental second-personal aspect of moral obligation cannot be reduced to non-second-personal relations or conditions. (Section III.B)

c) Second-personal relations (alone) cannot provide for the supremacy of moral obligations or an authoritative position of moral obligations within the deliberative standpoint. (Section IV.B)

If we accept the three preconditions discussed above, then we have a problem. We want to explain how there are moral obligations and they are supremely authoritative, but it seems that obligations must arise from something which cannot bring with it this practical normative authority. We seem to be left with making sense of moral obligations as normatively optional (giving up supremacy – but how would these be *obligations*?), or claiming that people simply *are* second-personally accountable and this just *is* normatively supreme (giving up an explanatory approach).[23] It looks like something has to go.

It may seem easy to just point out that moral obligations could derive from the joint action of first-personal and second-personal relations, where they each do separate work. Ultimately, I argue that this is true, but the way that it is true is not as simple as it may appear. I have argued that moral obligations must arise through relations of accountability between the obligated party and specific individuals, or the moral community at large insofar as they are, or it is in the right position to demand accountability. I argued further that these accountability relations cannot be reduced to first-personal practical rationality conditions.

[23] There is no real third option since supremacy without obligations means nothing.

These conceptual points about moral obligation seem to be rather firmly rooted. But then, there doesn't seem to be room for some first-personal mechanism to work alongside the second-personal accountability conditions. It is the demand itself (whether actual or hypothetical), in second-personal terms, which must give rise to the obligation. The normative authority of the obligation is supposed to arise through the demand. A successful model of moral obligation must meet the conditions of second-personal accountability *and* normative supremacy in full. The reply that the first-person and second-person standpoints work separately and alongside each other most obviously suggests the following two models for how obligation might work. The first is that someone demands A and A is (first-personally) rationally correct to do. The second is that someone demands A and this demand for A is (first-personally) rationally correct to follow. However, as I presently argue, neither of these models meets all of our stated conditions. They both would contradict the sense in which the second-personal relations that form the basis for obligation as accountability cannot be reduced to non-second-personal relations.

The first model fails immediately, because it relies on the rationality of A as independent of the demand for A. Say that you step on my foot and I demand that you remove it, while you also have reason to remove it because you might twist your ankle or because stepping on people's feet makes you uncomfortable. This does not provide for your accountability to me, and so does not provide for you having any moral obligation to remove your foot. It may seem that you have an obligation, but that is because of something that comes *from my demand* (or position to demand) that you remove it. This

model does not satisfy condition (a), that the obligation arises *through* second-personal reasons.

The second model is much more interesting, and much closer to what I, in the end, argue is right. The second model has it that in cases where the demand is rationally correct to follow, this gives rise to obligation. However, this also does not provide for obligation as accountability. Again, say that you step on my foot and I demand you remove it. On this second model you would have an obligation if you had reason (or better, conclusive reason) to follow my demand. This could follow simply from you having reason to be afraid of my wrath. Even a more refined version of this model would allow obligation if you had a reason to follow my demand because you thought I was respectable or because you took pity on me. But reasons of these varieties would be "reasons of the wrong kind" as Darwall puts it. They do not generate normative reasons from the second-person standpoint. They do not arise from the demand itself.

Another take on the problem comes from an objection by R. Jay Wallace to Darwall's presupposition argument (Wallace 2007, pp. 35-36). Wallace argues that Darwall himself fails to keep the normativity of accountability second-personal. In Darwall's attempt to uncover a deep pool of presuppositions to the second-person standpoint, he argues that addressees must accept second-personal demands with the authority of a common standpoint with the addresser – this is supposed to secure the common authority of the addressee. Wallace argues that this makes the authority of accountability non-second-personal. According to Wallace, the authority either comes from a purely first-person endorsement of the offered reason, or it makes no sense and is unexplainable on Darwall's theory (which would mean, of course, that Darwall had no

argument for it). I have already criticized this argument of Darwall's on other grounds, but Wallace's objection helps show the difficulty of trying to explain the normative authority of second-personal reasons.

In the next section I present a solution to the puzzle which works with my theory of practical normativity and is based on the idea that the source of the second-personal reason must be "buried deep enough," as I put it, so that it does not arise from any explicit first-personal reason. The problem with the two models suggested above is that the reason for performing the demanded action, the reason for removing your foot, derives from some explicit reason that you have to engage with the demand I make. Condition (a), that the obligation arise through second-personal relations, is not satisfied by either model, because they each rely on the work of a prior, non-second personal, explicit reason. This must be eliminated to create a successful model of obligation as accountability.

VI. The solution and resolution of the two conceptions of morality

A. General solution scheme

The problem I identified in the previous section is how to account for the normative supremacy (or an approximation thereof) of moral obligation within an explanatory system while bearing in mind that the bindingness of obligation must arise through purely second-personal relations. I further specified the problem with the claim that what is at issue is that second-personal reasons cannot derive from some other, non-second-personal reason. This condition is hard to meet, because pressure comes in the other direction to incorporate the force of second-personal reasons into a general scheme of deliberative practical normativity.

The solution, then, would be for there to be some rational force which is incorporated in the general deliberative scheme of normativity, but which applies to some second-personal relations without deriving from any other explicit practical reasons. That is, there must be second-personal reasons which do not derive from other explicit reasons at all, even though they can be incorporated into a general scheme of practical normativity. Of course, second-personal reasons could derive from other second-personal reasons, but that would not help us out of the problem, since explaining how second-personal reasons work at all would require leaving aside any reliance on further second-personal reasons. Also, second-personal reasons could, in a sense, be helped along by non-second-personal reasons, but not in any fundamental way, or they would not be fit to ground the accountability which is the basis of moral obligation.

Now, the picture I am outlining of how to solve the problem does avoid losing both the irreducible second-personal character of obligation and its normative supremacy, but it may seem to falter on the explainability requirement. If we have to make out a special variety of practical reason which is incorporated into a general scheme of practical normativity, but does not derive from any other reason of a generally first-personal kind, then we may seem to be left with no possibility of an explanation. However, as I see things, these are the conditions which have to be met for a full solution to the problem. I argue that my theory can account for all of these conditions in a fully explanatory way. Whether or not other theories could do the same is something I leave unanswered.

B. How my theory can satisfy these solution conditions

The conditions I have laid out for solving the problem of explaining normatively supreme, accountability-based moral obligations can actually be met rather simply on my

theory. A helpful concept for laying out this solution is *demand-receptivity*, the tendency a person has to be receptive to the second-personal demands of other individuals or the community at large. We can think of the problem as the question of what gives demand-receptivity authority from the individual's first-person, deliberative standpoint. Darwall did not successfully provide an explanation of this, but simply tried to show that a person was committed to some kind of demand receptivity from the second-person standpoint, which does nothing to account for its overall authority from the general deliberative standpoint.

Bringing in my theory of practical normativity, what gives authority to a particular practical principle, value or perspective from the deliberative standpoint is its place in the individual's systematic network of practical commitments. If I consider a practical principle like "I remove my foot from on top of someone else's if they demand it of me and I have no further reason to keep it there," this will have authority over my deliberation in accordance with how it would fit in with the rest of my commitments in terms of systematization. If it fits in more firmly and is better connected to the whole of my field of commitments than alternatives like "I do not heed the demands of others," or "I keep my foot on Bill's until he says 'uncle'," then it has authority and I have reason to act on it, and conclusive reason to act on it if there is nothing else at play.

Now consider principles which capture demand-receptivity in general. I might accept a principle like "I adhere to the demands of the community" or "I do what others are in a position to demand of me." Principles like these, or various normative stances which embody principles like these,[24] put demand-receptivity into the proper deliberative

[24] For a discussion of normative equivalence claims between practical principles and other ways to explicitly think about practical reasons and use the tools of deliberative normativity, see Chapter 3, Section

format. So, the issue of whether demand-receptivity has the deliberative authority we are looking for can be turned into the question of whether principles like these have systematic authority.

As I argued in Chapter 3, whether or not a practical principle has normative authority for an individual agent is a question of how it fits into that agent's systematic scheme of commitments. Various practical principles about what to do and when are accepted by an agent, either through explicit avowal or through the agent's dispositions and the implications of the agent's decisions. Some of these principles take a more general, theoretical form, and some take a form of prioritizing one's commitments. These kinds of principles help form a systematic scheme from the agent's overall field of practical commitments. Because the aim of deliberation is systematization, one always has reason to follow those commitments which do more to fill in the systematization of this overall scheme. Although hardly anyone thinks explicitly in terms of a systematic scheme, such a scheme is implied by an agent's dispositions and decisions – by the employment of practical reason. So, principles which increase or maintain systematicity have authority over principles which do not. Those principles or commitments which do the most work in systematizing one's field of commitments are those which have the most deliberative authority for the given agent.

Can we say that demand-receptivity has this systematic authority for us and for everyone we are concerned to account for? Here, our social nature, which Darwall points to in a misguided attempt to fill in his constructive arguments, can do the work. We are social creatures. We are, in fact, receptive to demands in many ways. This is a core aspect of our interaction with each other. Further, this is a guiding feature of how we

---

III.B.

relate to others and to the community as a whole. It shapes who we are, individually, as people who live in a social community with no definite boundaries. These observations suggest a special place for demand-receptivity.

It is true, on the one hand, that I cannot in any way formally demonstrate that demand-receptivity has authority for a particular individual, because a person's systematic scheme of practical commitments is just too complicated and too flexible for that. However, on the other hand, I take it to be highly plausible that demand-receptivity does enjoy a privileged position within our field of commitments, at least, for those we are concerned to account for. More importantly, an explanation of how demand-receptivity could have general deliberative authority is provided by this argument. It has authority just in case it enjoys a privileged position in one's field of commitments in terms of its centrality to systematization.

Two questions remain at this point: (1) Have I avoided grounding second-personal reasons in terms of other, non-second-personal reasons? (2) How can crucial terms like "in a position to demand of me" be understood? The second question can be answered by appeal to contractualism. This is exactly what Darwall does with his conclusion that we owe respect to each other as "free and equal." How to fill in what respect amounts to and what it is to acknowledge the authority of others is a large issue. In rough terms, however, we can say that this hole can be filled with the appropriate contractualist theory, which I explain more thoroughly in the next chapter.[25]

On the issue of avoiding a derivation of second-personal reasons from non-second-personal reasons, the beauty of constitutivism comes into full view. We can ground the deliberative authority of demand-receptivity directly in the governing of

[25] For more on this, see Ibid., , pp. 300-320**.**

deliberation by systematization. Thus, genuine authority is accounted for. At the same time, however, we avoid using further reasons, because *there is no normative reason to aim at systematization*.[26] The aim at systematization is constitutive of rational deliberation. It is the foundation of practical normativity. So, as I discussed in Chapter 2, there is no normative reason to conform to the constitutive aim, but the aim does govern deliberation and agency. The aim sets the standard for what rational deliberation is, and so sets the standard for what makes rational deliberation successful. Successful deliberation, according to Deontic Kantianism, also argued for in Chapter 2, is what makes for right action, and right acts attributable to agents in general. The aim governs deliberation and agency without having any practical normative backing. It governs deliberation and agency, because it is constitutive of what rational deliberation, and therewith rational agency are.

This means that on this model second-personal reasons are incorporated into a general scheme of practical normativity from the deliberative standpoint, and not via any other, non-second-personal practical reasons. When demand-receptivity has authority, it has in virtue of its role in systematization, but one does not have a normative practical reason to systematize one's commitments. The force of the authority comes from the constitutive aim, which determines what deliberative authority is in general, and constitutivism does not rely on any foundational practical normativity to account for deliberative authority. This is exactly the situation we needed in order to solve the problem.

[26] Of course, there may be some round about, derivative normative reason to aim at systematization in various ways in particular cases. But this is not what is at issue. It is the constitutive aim of deliberation which is not backed by practical normative reasons.

C. The whole picture – contingent, merely general normative supremacy of moral accountability

What I have laid out is a picture of moral obligation whereby obligation is conceived as stemming from accountability to other agents, as Darwall suggests. On my version, however, the normative supremacy of accountability is contingent and a mere empirical generality (not universal) among rational agents as we find them. Moral obligation does not spring directly from rational agency as Kant thought, but rather from the combination of rational agency and our contingent empirical social nature. Whether or not someone has moral obligations is as much a question of the individual's empirical social traits as it is a question of the individual's capacity to rationally deliberate and use practical reason.

This analysis leaves the conceptual question open as to whether there can be moral obligation which is not normatively supreme. I might have second-personal reasons to remove my foot, while these reasons are not conclusive. If I am accountable to you for removing my foot, does that mean that I have a moral obligation to remove it, but that it would be wrong for me to do so? I take this to be a purely conceptual issue about what obligation is and whether it can take this form. Those who take moral obligation to be conceptually supremely authoritative will say this is not obligation, while those who deny this may allow that there are such obligations. I will not attempt to answer this question, since I ultimately do not take it to have much import for my project.

The idea of contingent supremacy may be a bit of an odd one. It may elicit the objection that there can be no such thing. However, this is not so. Contingent supremacy is still supremacy. Having conclusive reason to do X does not imply that one would have had conclusive reason to do X had one been different. Because what we are considering

the supremacy of is distinct from the bare standard of correct choice, it makes perfectly good sense for there to be contingent factors which play into whether a consideration is conclusive or has supreme practical authority. In other words, while it is true that it would not make sense to talk of the contingent supremacy of the constitutive aim of deliberation – that must be conceptually true – it does make sense to talk of the contingent supremacy of other standards that get their authority from the constitutive aim. Derivative standards, like accountability, can get authority from the constitutive aim, systematization, and contingent empirical facts which feed into how things are systematized. It could be a contingent, but true fact that accountability for some individual is always supremely authoritative, because that individual's whole system of decision-making is centered around demand-receptivity in a way which cannot be undone. Maybe a person with an extremely acute social drive would be in such a position. For the rest of us, presumably, accountability will not *always* be supremely authoritative, because there will be other central principles which will have more authority in certain cases. For example, it may well be that for a lot of people saving one's mother from death has a more central role than does accountability to the moral community to save an airplane full of strangers. If this is the case, then for such people it will be right to save their mother instead of the airplane full of strangers. That might not mean that such people would be morally obligated to save their mother over the strangers, but rather it might mean that morality had no authority in such a case or that it did not apply, since morality (or at least moral obligation) is a matter of accountability. On the other hand, it could be that one is accountable to one's mother to save her, making

it a case of obligation, but this, again, is a complicated conceptual issue concerning the further specification of moral obligation.

Another objection that might be raised is the following. Darwall also seems to take up a view of moral obligation as contingent, although he does not acknowledge this very clearly. He argues that presuppositions of the second-person standpoint commit those who take it up to the equal authority of all. Then he argues that, even though there is apparently no conceptual or practical necessity in taking up the second-person standpoint, we cannot, as a matter of fact, avoid doing so. He does offer some other arguments about the metaethical acceptability of second-personal reasons, but I fail to see how this even suggests an explanation of the kind needed to avoid contingency. He seems to offer a contingent view of moral obligation. Now, if moral obligation winds up contingent on my view, how is this better than Darwall's own view?

The answer to this is that, even granting Darwall's presupposition arguments (which in fact I do not), he does not account for normative supremacy at all. Darwall does not even provide for contingent supremacy. He fails to give an account that puts second-personal reasons into a general normative scheme from the deliberative standpoint. Thus, Darwall does not provide for the deliberative authority of second-personal reasons in any way that would help us make sense of the normative force they are supposed to possess. Further, any theory that tries to explain moral obligation and its authority wholly from the second-person, or social standpoint, will run into the same problem. As my formal argument that the Supremacy Thesis must come from a general first person standpoint shows (Subsection IV.B), explaining the normative authority of

second-personal reasons must always be done in terms of a general, first-personal scheme of practical normativity. My theory, as I have argued, does all of this.

Chapter 5
Contingency, Subjectivism and the Moral Community

I. A complex picture of contingent practical life

A. The focus of this chapter

So far, I have argued for a theory of practical normativity and given an account, in the most general terms, of moral obligation. It may seem a bit confusing, however, how it is that I can claim to be giving a theory of morality. My theory may seem beset with a number of problems, insofar as it purports to provide an account of moral obligation. My theory may seem to make morality subjective in a number of senses. It may seem incapable of giving any genuine content to general rules for how we are to live in society. It may also seem not to provide any clear sense of what the relationship between moral obligation and cooperative living is supposed to be. The purpose of this chapter is to clear up these issues, remove the appearance of inadequacy and fill out the theory of morality I am putting forward.

There will be three primary stages to filling in my theory of moral obligation. The first two stages concern rejecting subjectivism in the forms that are connected to the particular nature of moral obligation, as opposed to practical rationality in general. The first addresses what the content and overall character of our moral obligations would be according to my theory. I reject the notions that my theory is egoist and that my theory does not provide the right kind of obligations. The second stage deals with formal constraints on a theory of morality. I investigate the relationship between moral

obligation and cooperative living. I argue that my theory provides for the right connections between how our social, public life relates to morality, in terms of moral debate and moral judgment.

In the third and final stage, I provide a sketch of the overall picture of morality I am advocating. I describe the important role that character – and so, in a sense, virtue – plays in my account of morality and right action. I argue that the distinction between what is practically normative in general and what moral obligation requires lines up with corresponding distinction between private practical life and public practical life. This makes room for the further claim that, in fact, my theory of morality supports a contractualist account of the content of moral obligations. In the end, I argue that our moral obligations come from the relations of accountability that we negotiate with each other through public discourse. I understand 'discourse' in a very broad sense that includes any form of communicative interaction.

B. Distinction between substantive and formal subjectivism

Perhaps the foremost problem with the theory of morality I have put forward in terms of its adequacy *as* a moral theory is its flirtation with subjectivism. I have outlined a theory which makes moral obligation contingent and merely general and allows for the content of moral obligations to vary with the groups and individuals for whom they arise. This seems to bring the threat of subjectivism to bear on the picture of morality suggested by my view. It is my task in this chapter to address this worry of subjectivism in the several ways in which it is most prominently suggested by these features of my theory.

We can distinguish two major forms of subjectivism which are relevant for us at this point in the discussion. While I addressed one major form of subjectivism in Part II

by arguing for the reality of practical normativity (which I argued in Part I is the most philosophically significant form of normative or moral subjectivism), here, I look at two forms of subjectivism that concern the content of obligations. This is the kind of subjectivism that is related to the contingency and lack of universality of moral obligations. We can distinguish between *substantive subjectivism*, which has it that moral obligations do not amount to the kinds of things we think they do concerning cruelty and kindness, inequity and fairness, selfishness and generosity and so on, on the one hand; and *formal subjectivism*, which has it that morality is too individuated and there is no sense to collective, public morality, on the other hand.

The broad idea of substantive subjectivism that moral obligations are not what we think they are or should be touches on many things. This clearly concerns an assumed general tendency of moral obligations toward not harming others and doing others service. Moral obligations are supposed to generally direct us not to harm others and, perhaps to a lesser extent, to do others service and to help others. Another general way this might be conceived is in terms of whether moral obligations direct us toward mutually beneficial cooperative living and away from egoistic selfishness, which is not beneficial to the community. Overall, substantive subjectivism is the theory or idea that morality does not have the right sort of content in reality – our conception of moral content does not match up with the reality of what content there truly is in our obligations and our normative system of right action.

Formal subjectivism also concerns the content of moral obligation, but it challenges our concept of morality in terms of the formal features of moral practice. Although other versions of formal subjectivism concerning the content of morality could

be conceived, I focus on how moral practice involves the collective treatment of moral norms. Since the contingency and mere generality of moral obligation, implied by my theory, is what gives rise to the looming possibility of subjectivism, I only discuss formal subjectivism that seems to follow from the lack of universality. What seems to follow from a lack of universality, in turn, is problems with the public treatment of morality. There are formal problems for the public treatment of morality in the sense that the point of the public discussion of morals and the attempt to come to common decisions on what to do or "what we owe to each other"[27], is put into question.

In this chapter I try to show that neither substantive nor formal subjectivism follows from my theory and my theory is on safe ground with the public and collective aspects of morality even though it implies that obligations are contingent and not universal. I first look at substantive subjectivism and I argue that my theory is not subjective in this way since it does not imply any relevant form of egoism, and it does imply, given our social nature, a general approximation of the Golden Rule and cooperative conduct. I then look at formal subjectivism and I argue that my theory is also not subjective in this way, since we can make perfectly good sense of the public discussion of morals and the collective attempt to work on moral problems within the terms of my theory. At the end of the chapter, I sketch an overview of the picture of morality I am putting forward.

[27] Here I quote Scanlon to invoke the relevance of contractualism for this issue T. M. Scanlon, *What we Owe to each Other* (Cambridge, Mass: Belknap Press of Harvard University Press, 1998)..

II. Substantive Subjectivism

A. Egoism

The first worry I want to address is that my view is egoist in a way contrary to what is acceptable for a moral theory. I would like to make clear at the outset that I do not mean to assume that it is in general a sufficient reply to egoism that egoism does not count as a moral theory because moral theories outline the way we are supposed to cooperate. The problem with such an objection to egoism is that it is irrelevant to deciding whether or not egoism has things right on its own terms. Egoism, as an action-guiding theory, is in contention for which action-guiding theory, if any, has things right. We may say that egoism is not a *moral* theory and so it is not in contention for the issue of which moral theory, if any, has things right. This question, however, is not that interesting, because we would surely want to settle at the same time whether there is any action-guiding system which gets things most nearly right.

However, I have already directly argued that my theory is an action-guiding system which gets things right. What is now at issue is whether and to what extent this counts as a positive solution to the problem of explaining morality. In other words, the question that remains is whether I have provided a *moral* theory. This naturally raises the issue of whether my theory falls under the category of egoistic theories which only recommend self-interested action.

There are two possibilities here, both of which I reject. The first is that my theory might be egoist in the usual sense, meaning that it recommends one pursue what is in one's own interest or what is in one's own *best* interest. There are many possible egoist

principles which might recommend promotion of immediate interest, broadest interest, interest over time, or greatest interest over time, all of which need supplemental accounts of what determines the course of action that meets the description. For example, we might think that one should always promote one's greatest self-interest over time, but it is not obvious how things should be weighted over time. Since that which brings me satisfaction now is a surer bet than satisfaction in twenty years, near interest seems to have an advantage over distant interest. There is also the question of when a person will die and no longer be able to benefit from past actions. These and many other complications must be addressed by a particular egoist theory. I argue that my theory is not egoist in any of these senses and that it does not imply any general normative egoist principle at all.

The second possible sense in which my theory might be egoist, is not by implying a general normative principle of self-interest, but rather in the instrumentalist sense, that my theory only directs one to take the necessary means to one's ends which are themselves beyond normative reproach. If my theory only implies that one should take the means to one's ends, then it surely would not count as a moral theory even when supplemented by the empirical claim that our social nature gives us the ends of promoting the general welfare and pursuing mutually beneficial cooperative living. This would not put people in the position that they had any binding duties to each other, because the directive to take the means to one's ends does not imply that one keep one's ends without rearranging them. If I have the end of being nice, and being nice in my present circumstances requires letting someone in front of me, instrumentalism does not imply I

should let the person in front of me rather than just changing or adjusting my end of being nice. Maybe I should have the end of being nice when I feel like it.

Each of these two possible worries for my theory can be set aside fairly easily. I can separate my view from traditional egoism, by appeal to the fact that I have given a general account of practical normativity. Just as we can make a distinction between second-personal and non-second-personal reasons, so also can we make a distinction between egoistic and non-egoistic reasons. *Egoistic reasons* are reasons deriving from subsidiary aims or practical principles with a view toward promoting one's self interest. If a practical principle is formulated as aimed at serving one's own interest, it is egoistic. If not, it is non-egoistic. There is no general egoistic principle providing for the authority of egoistic reasons implied by my theory. My theory does not imply egoism or the authority of egoistic reasons, because non-egoistic reasons can arise from the aim at systematization and one's motivational dispositions just as much as egoistic reasons.

The instrumentalist worry is even easier to dismiss. The problem with instrumentalism is that it provides no stability for what one should do. There is never any practical 'must' that arises from instrumentalist directives to take the means to one's ends, because instrumentalism provides no guide to the infinite possibilities of shuffling and reshuffling one's ends in relation to available means. My theory provides this stability. The constitutive aim at systematization of one's commitments creates a well-structured standard of authority among one's ends or commitments. One's ends are not left free and open, because they must be regimented by systematization in terms of the standards of practical normativity. Whether or not I should be nice in the sense of letting some person in front of me, is decided by whether or not doing so requires acting on a

practical principle which has authority over alternative practical principles (which do not direct one to let the person in front) in terms of systematization.

B. The Golden Rule

In order to reject substantive subjectivism, it is probably less important to dismiss egoism than it is to establish some positive rules that resemble our typical conception of what is right and what is wrong. In what follows, I argue that an accurate empirical generalization of some of some of our obligations is captured by Darwall's second-personal interpretation of the Golden Rule. Versions of the Golden Rule are probably the most culturally and historically universal of general practical principles, and therefore, establishing a worthy version of it is about the best way to refute substantive subjectivism.

As a caveat before beginning my argument from systematization and our social nature to an authoritative, general Golden Rule, I would like to make some ancillary points about deriving general rules from my system and the specific nature of the Golden Rule. The first point to make is that we cannot expect a precise and rigorous derivation of any generalized principle, because it necessarily relies on loose generalizations and enthymemes. Since the authority of practical principles comes from its position within the entire scheme of commitments for an individual, we must take a majority of the systematization as background whenever we consider the authority of principles, because it is pragmatically impossible to consider the scheme's systematization all at once. So, we must start with various "given" practical principles and normative stances, although we will need to be flexible in our treatment of them. Further, the problem is compounded when we look for generalizations that hold for most or all of a group of people, because

we then have to start with generalizations of practical principles and then work flexibly with those.

Nevertheless, the project is not impossible. We can look at various characteristics, which we can agree to be common, and make plausible their authoritative position within people's schemes of commitments, and then show how they jointly lead to some practical principle which can operate as a generalized rule. This procedure leads to empirical generalizations and rules of thumb. The result of the investigation is a rough guideline for what is authoritative for a group of agents, where the more firmly founded the authority of the principle, the looser the generalization and the vaguer the application. Nevertheless, generalized authoritative principles can be found by this procedure.

The second point I want to make concerns the specific nature of the Golden Rule. As I just noted, a high position of authority and a well-generalized character of a principle will correspond to difficulties in its application. Similarly, it is significant that the Golden Rule has limited scope. It does not really compete to be an absolute principle determining all actions. So, what is at stake in the ensuing argument is not a master principle of action for all moral agents. There are many situations in which the Golden Rule will not tell the agent what do, even in conjunction with reasonable supporting principles. For example, the Golden Rule does not tell us how to dole out charity (except perhaps in some limited sense), whether abortion is right, how to treat non-agents, and, in general, does not help in cases that are not primarily interactive. It is, however, a helpful start to seeing how our social nature supports moral obligations of the commonly championed varieties.

The argument for the Golden Rule that I want to make is very simple, and involves no complicated analysis of our specific commitments. The idea is simply that our social nature, and specifically our commitment to the second-person standpoint, give authority to second-personal principles that help us organize our social lives through a more equality-oriented stance. Darwall's second-personal interpretation of the Golden Rule is as follows. "One should not act in ways that one demands or expects (or would demand or expect) others not act, or equivalently, in ways that one would resent or object to" (Darwall 2006, p. 117). Darwall argues that this is, intuitively, the best way to understand the Golden Rule on its own, regardless of how it helps us to derive it (Darwall 2006, pp. 115-117).

However, we can see that this way of understanding it makes clearer the source of its authority. For people living in society, the goal of living cooperatively through a system of accountability to one another is highly authoritative. This is the authority of the second-person standpoint. I discussed this authority in Chapter 4 and pointed out how it comes, not directly from presuppositions of second-personal address as Darwall holds, but from contingent facts about our social nature and our aim at systematizing them as rational agents. In other words, we are predisposed in such a way that we are committed to morality, which is a system of accountability relations that function to govern cooperative living.

To put this in context, let us note some generalities and see how they relate to each other. We have various ways of acting and various desires for how we should be treated. We also have various concerns of sympathy for others, as well as second-personal attitudes and dispositions to second-personal attitudes and stances like

resentment, gratitude and guilt. While these attitudes and dispositions do not directly imply that we commit myself to respecting others, they do create pressure. The pressure comes from the difficulty of systematizing our approach to the second-person standpoint, our dispositions to second-personal attitudes, and our common goal of living cooperatively with others through a system of accountability and second-personal relations, without accepting some sort of equality of accountability principle. This pressure of systematicity is best alleviated by subscribing to a Golden Rule-like principle which organizes many of these central attitudes concerning how we act and how we interact with others. Because the Golden Rule does so much organizing work – that is, so much systematizing work – it enjoys great authority among our commitments. For a normal, socialized member of society, it is nearly impossible to avoid high-level commitment to the Golden Rule, because there is no equally good, non-equivalent way of handling these practical pressures.

This means that substantive subjectivism, of the kind to which my theory might seem to lead, can be set aside as a barrier to my theory vindicating morality. My theory does imply that we have the kinds of obligations that we think we do, at least in broad strokes, even if not in finer detail. The Golden Rule captures much in the way of injunctions against harming others, especially without reason, and even gets at the idea that the needy should not be left completely abandoned, even if it does not help us settle how and when to give charity, or how much service we owe to our fellows. My theory does not imply selfish egoism, or callous instrumentalism, but instead provides for the general authority of a kind of equality of accountability.

III. Formal Subjectivism

A. Applied ethics and morals in public discourse

As I said at the beginning of the chapter, there may seem to be certain formal problems with moral practice that may seem to derive from the contingency and mere generality of moral obligations. The problems of formal subjectivism with which we should be concerned center mostly around the role that the public discussion and recommendation of morals is supposed to play in our moral practice. It seems essential to any purported vindicatory interpretation of morality that it be able to make sense of what the point of discussing morality in a recommendatory vain is supposed to be. If we are engaged in applied ethics, that is, the investigation of what *we* should do in various situations, there should be some sense to what we are trying to achieve in so doing. Applied ethics, or the informal discussion of morals should serve a greater purpose than merely providing emotional therapy for the participants. When I argue to the public that abortion should be legal and that it is perfectly moral across a broad range of circumstances, I should be doing more than just venting my frustration or making myself feel important. I should, in some sense, be sensibly trying to show that a relatively open policy on abortion is what *we* should have.

Thus, if I am really putting forward a moral theory, I need to be able to show that, on my view, there is a perfectly good point to doing applied ethics or to engaging informally in the collective determination of moral policies and practices. It is more or less an unacceptable result of a theory of morality, unless it is to be an error theory, that it can make no sense of the point of public engagement in moral discussion. A theory of morality must be able to handle the sense in which we try to persuade one another,

through public discourse, that we should be pursuing a certain policy, and that we can succeed in this. The arguments used in such discussion are supposed to be aimed at showing that certain policies are what we should adopt and live by, and these arguments, at least in principle, should be capable of success. It would be unacceptable for a non-error theory of morality to not be able to make sense of how we can succeed in showing that various policies of practical life are what we should live by. If, however, I can show how public discussion of morals does make sense on my view, and has a purpose that fits with what we think we are doing when we engage in such activity, then I can avoid the problem and defuse a potentially powerful objection.

A further issue of looming formal subjectivism concerns the status and justification of our reactions to others behavior in terms of moral judgments and holding people responsible. It may seem problematic that we should impose judgments on others and hold them responsible for doing or not doing various things when what is right for a given individual depends on so many things internal to that individual's motivational and principled makeup. Saying that a man was wrong and should be punished for abusing his child might seem misplaced if we consider that there are no completely general rules about child-abuse being normatively prohibited for rational agents. The man might not be a rational agent in my refined sense, or he might be a rational agent, but not have the other empirical traits which would jointly make it a wrong act or action to abuse one's child. This seems to put pressure on my theory in the sense that my theory seems to undermine our legitimacy in judging others by the standards of accountability we endorse and in holding them responsible, whether through punishment or otherwise. This

judgment problem is also a form of a formal subjectivism problem, and is a problem I need to explain away.

B. Two directions to systematization – public to private and private to public

It can be seen that there is perfectly good sense to the public discussion of moral policies and practices in a recommendatory vain. Of course, on my view, every practical purpose, if it is to have normative significance, must, ultimately, be grounded in systematization. So, I must argue that whatever purpose this practice has comes as a form of systematization. In fact, I outline two related and reciprocal purposes to the practice of doing applied ethics and informal ethical discussion. The systematization that is achieved by public discussion of morals goes in two directions: individual to group, and group to individual. The participants work to jointly systematize the group's commitments taken as a whole, and the group helps the individuals to systematize their own commitments by sharing views and ideas.

As I elaborate on these claims in what follows, I present a somewhat simplified version of what is under examination. In this section, I am not concerned to analyze how the differences between right action for an individual in general and moral obligation relate to the systematization produced from individuals for the group and from the group for individuals. The way in which moral obligation stands out as a special kind of practically normative standard (stemming from accountability relations) has a special relationship to the systematization we try to achieve for the group. However, this is a topic I take up in the final section. When I do take that topic up, I discuss how contractualism plays a role as a way to understand what is going on when we work to systematize the group's commitments in order to structure our accountability relations to

one another. Contractualism can be thought of as an account of how systematization for the group governs moral obligations. In the final section, I sketch a picture of this interplay.

The first sense in which public discussion of morals works to help systematize commitments is that we work to establish common commitments of the group. In order to achieve the overall common goal of cooperative living, we must attempt to align our commitments with one another. To do this, we try to set out commitments we share in common, which belong to all of us in an active moral community. In fact, what puts us in a *moral community* together is exactly this activity, trying to establish common commitments (and accountability relations in particular). When I try to persuade you that abortion is often acceptable, the rich should be taxed at a much higher rate, one should deal harshly with sex offenders, or any sort of "moral" point, I am trying to get us to agree on a common policy. Of course, the policy I push for is probably the one that most closely springs from what I take myself to be committed to. Discussing morals in this way, whether through casual conversation, published applied ethics papers, popular media discussions, or aggressive arguments on the street, is a method for creating common commitments among a moral community. Cooperative living depends on us being able to coordinate our practical principles and rely on each other to live by roughly the same rules, or at least the same general structure of rules. Maybe we grant the king certain privileges, but we mutually work within the system of rules that grants the king a special status, and this is openly part of our common commitments. If not, then we are not in a common moral community, or at least not a well-functioning one.

The primary purpose of the public discussion of morals is to bring the community to a common standpoint of practical life. This is true according to a default conception of moral debate, whereby the public discussion chases after absolute rules or general guidelines that hold for everyone, regardless of what they are like beyond being theoretically rational and capable of consciously controlled behavior. On this conception, the point of arguing that cruelty to animals is wrong is to show that there is wrongness to it for anyone, but it is also to show that everyone should agree to accept and live by the idea that cruelty to animals is wrong. Showing that there is an independent fact of the matter, which does not depend on a person's internal motivational forces, is a means to the end of showing that we should all accept and live by the practical principle that we not be cruel to animals. This ultimate purpose to moral debate of showing that we should accept and live by various practical principles is perfectly preserved on my view. That is the primary purpose of the public discussion of morals.

There is also a further, reciprocal purpose to public moral discussion. We also work to help each other systematize each of our own commitments. By sharing views and ideas with each other, we not only get each other to change our own individual way of organizing our commitments in order to meet common goals, we also help each other organize our commitments without reference to common goals. This is how the group helps individuals systematize. The group to individual direction of systematization assistance works simply through sharing ideas. I try to systematize my commitments – that is, I try to maintain and improve my systematization as problems with it are presented by practical dilemmas. Talking to you might be a good way to figure out what I should do. But, this does not have to take the basic form of coming to you for advice

and asking you what you think I should do in a given situation. It might also take the form of a general discussion about what policies are right to pursue. Done in this more general, and more abstract way, we can help each other at the same time, and work on a steady or recurring problem all at once. If you and I are both ethical vegetarians who try to abstain from supporting commercialized factory farming, we might find it perplexing whether and how much we should use leather products. If we discuss with each other what the right thing to do with regard to this tension is, we can help each other find the best systematization of each of our sets of commitments.

These two directions of public work toward systematization bleed into each other, because in working on systematizing the group's commitments we are working on systematizing our own commitments that concern social interaction. Of course, if we come to a common way for us to systematize our commitments, then we must do so by setting out a way for each of us to systematize our own commitments. Analogously, if we're standing around in a circle and we come to a common spot to huddle together, we must do so by each of us taking up a spot. It is possible that one or more of us does not move in order to form the huddle. The huddle does not have to happen in the middle of the circle. However, each of us must take up a spot, whether we were already there or not. In terms of systematization, if we find a common way to organize our principles as they concern a set of practical issues, then each of us must take up that way of organizing in order for us all to accept and live by the proposed common commitments, whether we already embraced them or not. The rules we develop must, in some sense, work for everyone as principles which help systematize and provide an organizational scheme.

Similarly, in systematizing our own commitments in a group context, we work to bring our own commitments more in line with a general system of group commitments. If we try together, as vegetarians to work on the use-of-leather problem, we naturally look for a common solution. Theoretically, the right solution for each of us could be different, but, however likely that may be, this problem in isolation will probably concern working out the different stances that could be taken and figuring out how to smooth them over with other, related principles. It is more work to do this twice. To go back to our illustration, if we are working on how to treat leather we might consider the following two options: first, we might take any use at all of animal products deriving from factory farms to support factory farming and so to be wrong, or, second, we might take leather to be a mere byproduct of animal production (driven primarily by meat) and so to be irrelevant to abstaining from support of factory farms. What we need to do to decide between these options is see which one runs into more trouble with other practical principles (many of which will call for a review of empirical facts). As we examine these relationships, we work on making everything systematic (roughly, consistent and unified). We may run into more trouble with one of the options, which would lead us to work on the other. Hopefully, we will eventually reach a point where we seem to have worked one of the options into our system of principles, while the other will be left problematic. If we treat our respective schemes of systematization as roughly the same with respect to this problem, our work will be much easier and more effective. This means that the best way to help each other work on each of our own systematization regarding a common problem is to work toward a common solution. Thus, the two directions of systematization from the public discussion of morals run together.

C. Moral judgment of others

The problem of justifying and making sense of judging others and holding them responsible in a context in which their own systematization may not lead to the same standards of conduct as we have, needs its own solution. However, this solution is much simpler and closer to hand than the solution to the problem of public discussion of morals. The core of the solution springs directly from the fact that holding people responsible is an act. As such, this act, the act of holding people responsible, falls under the standards of systematization like any other act. Similarly, judgment of others, when regarded as a practical act, also is justified by the standards of systematization. There is a complication with judgment, since judgment can come in the form of a strictly theoretical exercise, when we are being "philosophical," and then different standards of correctness are at play. In the theoretically-minded case, when a judgment is aimed at being true, then the standards discussed in Part II concerning systematization and the added layer of the authority of social accountability for the acting agent are what determine the truth of such judgments. Moral and normative judgments are "agent-relative" but not "judgment-relative" because the truth of such judgments depends on the particular standards of authority of the acting agent, no matter what perspective of judgment is taken. In general, however, we can say that the legitimacy of responses to others' behavior is a question of our own systematization, because it is a practical, not a theoretical question.

Holding others responsible is an act. More specifically, it is a way to treat others. How we treat others, is a matter that should be decided from our own systematization of principles, which includes the way we systematize principles together, as a group. We as a group, and we as individuals must decide how to treat people, whether they be within

our moral community or outside of it. How we hold people responsible is part of this. We as a group, and we as individuals, must decide how and when to hold people responsible. We can take into account the particular motivational forces and reasoning of the agents we propose to hold responsible, or not. We can take the internal forces of agents into account to any degree, but this must all be decided as a question of how *we* should act. Holding people responsible is a way of acting, and so deciding when and how to hold people responsible must be decided on the terms that are appropriate for deciding how to act in general. Together, and as individuals, we decide how to respond to the acts of others and when and how to hold them responsible. We probably should take internal motivational forces into account to some extent, but not entirely. Indeed, this is what we already do. We allow some leeway in holding people responsible and punishing them, based on their internal forces, but we do not allow this to completely determine whether a person should be *held* responsible. To what precise degree this should be done, is a complicated and independent question from the issue of what grounds we have for holding people responsible at all.

Similarly, judging others can be done, or at least can be regarded as a practical act. This is especially clear if we think of judging others as making a public declaration (to our friends and neighbors, or maybe just to ourselves). I can declare that what she did was wrong and what he did was right, and these are practical acts. As such, they fall under the standards of practical normativity (assuming they are acts attributable to me), which are the standards of systematization of my commitments. I should perform these acts or actions just in case the systematicity of my commitments bids me to. Presumably, being honest and sincere about such matters will rank highly among my commitments

and will carry a lot of weight concerning what kind of judgments of others' behavior I should make. But, such declarations are acts and so they are ultimately right or wrong in accordance with the deliberative aim at systematization.

Unlike holding people responsible, judging people can lie strictly in the theoretical arena, when we are being "philosophical" about people's behavior. Philosophically speaking, a judgment that someone's behavior was wrong is true just in case it failed to conform with the aim at systematization of the acting agent's commitments. This is a theoretical point. What makes a judgment about the rightness of someone's behavior true has to do with what is the case for the acting agent. It does not matter what is the case for the judging party. This theory is what I call "agent-relativistic," that is, rightness varies across agents, but it is not what I call "judgment-relativistic," that is, it is not true that rightness varies according to the judge from whose perspective the act is assessed.

Theoretically judging behavior is extremely complicated, because there is so much to consider in determining what systematization bids. In practical terms, we arrive at standards that help to simplify the field of possible right and wrong actions through joint discussion (and other methods, perhaps non-democratic ones) of morals. This process helps, not just to figure out theoretically how we must systematize, but to practically provide ways of systematizing – to give ourselves ways of organizing our commitments. On the theoretical side, however, we must try to figure out what is already there, and to get things exactly right.

Simply put, it follows from the present account of right action and moral obligation that ethics is for *us* in the sense that ethics is a procedure for *choosing* which

principles to live by. Thus, how we apply our own moral principles to others is just another ethical matter. It is not metaphysical or epistemological in any deep sense. We may be concerned about how we hurt others, by forcing them into our moral framework (because our moral principles tell us not to hurt others), but we needn't worry about the absurdity of applying our moral concepts to others whose sentiments may or may not differ from ours. We do not need to get an exact theoretical analysis of someone's behavior before we begin to consider how to deal with the person in practical terms. Our standards for dealing with people come from the practical realm of policies that work for us (in terms of systematization). These policies will probably not require any exact theoretical analysis of what is right for, say, the north African who wants to mutilate his daughter's genitals, before we can figure out that we should hold him responsible for refraining from doing this. The degree of theoretical analysis required for a legitimate act of holding someone responsible (or judging the person in the practical sense) is a practically normative question.

IV. The final picture

A. Character and virtue

In this final section, I want to sketch the general picture of practical life I am advocating, especially as it pertains to the discussion of this chapter. I say a few things about the tension between the public side to practical life and the private side, how this lines up with moral obligation and right action, and where contractualism and normative moral theory come in. First, however, I entertain a brief discussion of how character and

character-development fit into my picture of practical life, which is, in turn, a discussion of virtue.

Throughout Parts II and III of this essay, I focused attention on the role that rational deliberation plays in practical normativity, and how that relationship leads to moral obligation. However, most of our intentional behavior, or at least a great deal of it is done, not from rational deliberation directly, but from character. This I mean in the sense that one reacts to one's circumstances through the use of non-rational cognition, which is partly trained by rational choices. When I sit down in a chair, I usually do it intentionally, but I do not usually do it through, or as a result of rational deliberation. Instead, there are various non-rational cognitive processes that take charge in bringing me to sit and carrying out the sitting. This, in general, would count as what I called in Part II an act attributable to me. It counts as an act attributable to me, because, even though sitting in the chair was not a rational action for me in the imagined case, it was brought about through dispositions of mine which were trained by me through rational decisions. I make explicit decisions from time to time which affect my character as it pertains to deciding whether to sit. This character training, in part, leads to me sitting in certain situations. In such situations, when I sit as a result of that character training, my act of sitting is attributable to me, because it is attributable to past rational decisions of mine. If my act of sitting is in no way the product of rational decisions, then it is in no way attributable to me (despite the use of the phrase "my act of sitting").

These conceptual points about acts being attributable to agents when they are not the direct result of rational deliberation all come from my arguments in Part II for the Attribution Thesis. The point I want to make here concerns the central role this plays in

practical life. As I said above, perhaps most of our intentional behavior is done from character. This role of character can run very deep, too, since character stemming from rational decisions, which develops more character, passes on its attributability to the agent, so that one's character can really take on powerful place in one's practical life.

This, in turn, means that developing one's character is of central importance. Since most acts attributable to a given agent come from the agent's character, the agent's character has a profound impact on the agent's success at systematization. Say that I train myself to be well-mannered. Acting and reacting in a well-mannered way, as a result of the training, can come up far more than the initial training actions. The developed character has a far-reaching effect. Its effect reaches farther than an isolated decision to be polite or to be rude to someone who has annoyed me. Decisions affecting character can have an amplified relevance to one's principles, since acts being in line with principles concerns what outcomes the principles target and what outcomes the acts bring about or tend to bring about. The impact of decisions on character is especially weighty, and so they must take a central role in decision-making.

Furthermore, this aspect of practical normativity is virtue-oriented in a fairly strong sense: developing character is important, not because of the consequences and not because it leads to right actions or good intentions. It is important because doing so is practically normative for the individual. Of course, this is not fully a virtue theory, but it gives a prominent and central place to character and character development. More specifically, developing character is important, not just because it leads to good future decisions by improving one's ability to systematize commitments later on. It is also significant for systematizing one's principles at the time of character development. If,

again, I train myself to be well-mannered, not only do I put myself in position to make rational decisions later that are in tune with being well-mannered (and all that stands behind that). I also work to uphold the forces that support being well-mannered now, by setting myself up to act well-mannered at later times when I make no rational decision. I set myself up to react, without rational deliberation, in ways that are in line with my current system of principles. The point of the training is not just to help make decisions later; it is also a good decision *now* to make myself well-trained – trained in a way that agrees with my system of commitments. Having a good character, in a sense, is good because it produces acts which are in line with who the agent "wants" to be.

B. Right action versus moral obligation and contractualism

In this final subsection, I review the differing roles of the practically normative in general and moral obligation. In Chapter 4 I laid out an account of moral obligation according to which obligation is what we are bound to do by the normativity of being accountable to the moral community. The moral community is the group of mutually accountable rational agents with whom one shares accountability. The bindingness of moral obligations must arise through the demands, real or hypothetical of particular members of the moral community or the community as a whole. These demands are themselves normative, because the general principle of being a member of the moral community enjoys authority within the systematic scheme of commitments of the individual agent. Thus, the normative supremacy of moral obligations is contingent and merely general among rational agents. There could be, and surely are rational agents who are not members of the moral community, and so have no moral obligations, because, even though they aim at systematization of their commitments in deliberation, they do not

have the motivational dispositions and organizational structure which make participation in morality authoritative for them. The existence of such individuals, however, does nothing to undermine the authority of moral obligation for the majority of us. For most humans, participation in morality is authoritative (presumably), because our social nature and personal needs make deference to morality the only way to successfully systematize our practical principles.

Since there is this definitional gap between what a rational agent should do in general and what moral obligation requires, the question arises as to how this plays into our general picture of practical life. Also, how this divide relates to the public discussion of morals is of particular interest. In Section III, I discussed the role that the public discussion of morals plays in practical life. I did not, however, distinguish there between the way that public discussion pertains to moral obligation as opposed to other practically normative areas or the practically normative in general. In order to get a clearer idea of where morality sits within practical life, a discussion of the differing points of focus between right action in general and moral action in particular is needed.

As a side point before moving on, I briefly dismiss an annoying objection which might be raised. One might object to my drawing a conceptual distinction between 'right action' and 'moral action.' Some might contend, that 'right action' is conceptually equivalent to 'moral action' or some such thing. However, while there is certainly a strain of the ordinary language use of 'right action' that is tied in this way to 'moral action,' as in the phrase "that action was not right," where the speaker is making a moral judgment, there is also a strain that is conceptually distinct, as is supported by the following argument which does justice to all of the concepts at least partially.

1. Right action is what one should do.
2. What one should do is what one has most reason to do.
3. What one has most reason to do is what is unqualifiedly normative for the agent.[28]
4. So, right action is what is unqualifiedly normative for the agent.

I define 'right action' in terms of what is unqualifiedly normative. So, I am following a legitimate strain of the ordinary meaning of 'right action', which here differs from 'moral action.' Furthermore, terms should be used as they are most effective in conveying a point or helping to increase our understanding of whatever is under investigation. Since I have defined a difference between 'right action' and 'moral action' which I have shown to be an important difference for our practical lives, I am justified in using the terms in this way.[29]

The basic dividing line in practical life between right action and moral action runs along an individual-group partition. Practically, it is right action which is ultimately of consequence and is the most important to track for the individual actor. This is so by definition. Right action just is action which is practically normative (supremely) for a given agent. The individual agent, necessarily, should be most concerned with where right action lies, and what is, overall, right for that agent to do. So, as the agent thinks about practical life, right action in general is what should be thought about and kept track of. The agent should not get overly caught up in thinking about some particular type of reason for acting, without regard for that reason's position in terms of right action all

[28] For discussion of 'unqualifiedly normative', see Stephen Darwall, "Internalism and Agency (Volume 6: Ethics)," *Nous-Supplement: Philosophical Perspectives* 6 (1992), no pages..

[29] For more on the legitimacy of using terms as they are most effective in increasing understanding, see the discussion in Chapter 4, Section III.B and Carl Hammer, "Explication, Explanation and History," *History and Theory: Studies in the Philosophy of History* 47, no. 2 (2008), 183-199..

together. If I think solely about moral action, without considering whether what I am thinking about is also right action (*de re*), then I am erring as a rational agent.

However, this does not mean that moral action is a useless concept or that it can be legitimately ignored. The primary reason for this is that for public discussion, it is moral obligation that is the most important to track. This is what we are concerned to hone in on with each other. As I discuss below, moral action is what is most important for the group, the moral community at large, and various public groups as particular units, to settle on. As I argued in Section III, the public discussion of morals does the work of structuring how we are to interact with one another and to facilitate the common goal of cooperative living. The primary standard at stake in such discussion is moral obligation, because it is our relations of accountability to each other that govern our common standards of cooperative living.

Since I have, in some ways, taken on Stephen Darwall's notion of moral obligation, I will also, in some ways, take on his idea that there is an important relationship between moral obligation as accountability and contractualist moral theories. The relationship that I want to point to is that there is a special connection between the public discussion of morals and of what we should do, on the one hand, and accountability relations on the other. One way to understand what contractualist theories are about is to take them to hold that morality is about agreement, hypothetical or actual, which provides the content of moral obligation. I have already argued that the content of morality comes from our accountability relations to one another. This is a social approach to morality, which sets it up to be allied with contractualism, which is also a social approach to morality. If we put these ideas together, we get the hybrid idea that

our accountability relations are filled in, or structured by our agreements. This is where contractualism comes in to provide an important component in the moral theory I am arguing for.

While on the one hand we would need to look at particular contractualist theories to see how best to finish the account of structuring accountability relations, we can make some basic points, which have already been raised by the idea that public discussion concerns moral action. In broad terms, the effect of the public discussion of morals is to structure our accountability relations with each other. I have argued that the public discussion of morals serves to help us systematize our collective commitments and bring them in line together. I have also argued that the primary focus of public discussion of right action is moral action, which implies accountability relations since they are the springboard of moral obligation.

Now, the basic contractualist idea is that some form of agreement is the basis of shared moral obligations. Since the public discussion of moral obligation within a moral community (or *the* moral community) is a way of creating a common authority of socially derived standards of conduct – accountability relations – by way of a deep form of agreement – organizing one's principles around common schemes of systematization – public discussion of moral obligation is a contractualist source of moral obligation. We create our obligations through agreement by discussing them and making them authoritative through shared adoption of systematic schemes and principles.

This process is how we come to have shared standards of accountability. By openly negotiating with each other what our shared standards should be, we come to arrangements which work for us in terms of systematizing commonly held commitments.

I have various commitments concerning interaction with others and participation in social life. Everyone else (in the moral community) has various commitments concerning interaction with others and participation in social life. Many of these commitments are already held in common. Some of them are not but could be replaced by commitments which were held in common. The underlying assumption of membership in the moral community is that there is a way to set up a common scheme of systematization among the community concerning participation and interaction. Open negotiation is the way to work toward such a common scheme. The result of this negotiation is structuring our common accountability relations. In this way, moral obligations stem from our commitment to society and the social project of structuring accountability through public discussion and negotiation.

Bibliography

Ainslie, G. (2001). *Breakdown of will*. Cambridge: Cambridge University Press.

Audi, R. (1997). Moral judgment and reasons for action. In G. Cullity, & B. Gaut (Eds.), *Ethics and practical reason*. Oxford: Oxford University Press.

Baumrin, B. H. (Ed.). (1969). *Hobbes's leviathan: Interpretation and criticism*. Belmont, CA: Wadsworth.

Baumrin, S. B. (2004). The shoes of the other. *Philosophical Forum, 35*(4), 397-410.

Baumrin, S. B. (2006). Becoming moral. *The Philosophical Forum, 37*(3), 321-332.

Boyd, R. (1988). How to be a moral realist. In G. Sayre-McCord (Ed.), *Essays on moral realism*. Ithaca: Cornell University Press.

Brandt, R. B. (1996). *Facts, values, and morality*. Cambridge: Cambridge University Press.

Brandt, R. B. (1979). *A theory of the good and the right*. London: Oxford University Press.

Bratman, M. E. (1998). The sources of normativity. *Philosophy and Phenomenological Research, 58*(3), 699-709.

Bratman, M. E. (1999). *Faces of intention: Selected essays on intention and agency*

Bratman, M. E. (2003). A desire of one's own. *Journal of Philosophy, 100*(5), 221-242.

Bratman, M. E. (2007). *Structures of agency: Essays*. Oxford: Oxford University Press.

Brewer, T. (2002). Maxims and virtues. *Philosophical Review, 111*(4), 539-572.

Brink, D. (1997). Kantian rationalism: Inescapability, authority, and supremacy. In G. Cullity, & B. Gaut (Eds.), *Ethics and practical reason*. Oxford: Oxford University Press.

Butler, J. (1726,1983). In Darwall S. (Ed.), *Five sermons preached at rolls chapel* Hackett.

Chisholm, R. (1976). *Person and object : A metaphysical study* Open Court Pub. co.

Colter, L. W. (1998). The sources of normativity. *Review of Metaphysics, 51*(4), 940-942.

Cullity, G., & Gaut, B. (Eds.). (1997). *Ethics and practical reason*. Oxford: Oxford University Press.

Dancy, J. (2000). *Practical reality*. New York: Oxford University Press.

Dancy, J. (2004). *Ethics without principles*. Oxford, Clarendon Press: Oxford University Press.

Daniels, N. (1979). Wide reflective equilibrium and theory acceptance in ethics. *Journal of Philosophy, 76*, 256-282.

Darwall, S. (1983). *Impartial reason*. Ithaca: Cornell University Press.

Darwall, S. (1985). Kantian practical reason defended. *Ethics: An International Journal of Social, Political, and Legal Philosophy, 96*, 89-99.

Darwall, S. (1989). Motive and obligation in the british moralists. *Social Philosophy and Policy, 7*(1), 133-150.

Darwall, S. (1990). Autonomist internalism and the justification of morals. *Noûs, 24*(2), 257-267.

Darwall, S. (1992). Internalism and agency (volume 6: Ethics). *Nous-Supplement: Philosophical Perspectives, 6*

Darwall, S. (1995). *The british moralists and the internal ought': 1640-1740*. New York: Cambridge University Press.

Darwall, S. (2002). *Ethical intuitionism and the motivation problem*. Oxford: Clarendon Press.

Darwall, S. (2006). *The second-person standpoint: Morality, respect and accountability*. Cambridge: Harvard University Press.

Darwall, S. (2007). Moral obligation and accountability. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics, volume 2* (pp. 111-132). Oxford: Oxford University Press.

Darwall, S. (2007). Reply to korsgaard, wallace, and watson. *Ethics: An International Journal of Social, Political, and Legal Philosophy, 118*(1), 52-69.

Darwall, S., Gibbard, A., & Railton, P. (1992). Toward fin de siecle ethics: Some trends. *Philosophical Review, 101*(1), 115-189.

Darwall, S., Gibbard, A., & Railton, P. (Eds.). (1997). *Moral discourse and practice: Some philosophical approaches*. New York: Oxford University Press.

Davies, P. S. (2007). What kind of agent are we? A naturalistic framework for the study of human agency. In D. Ross, D. Spurrett, H. Kincaid & G. L. Stephens (Eds.), *Distributed cognition and the will: Individual volition and social context* (pp. 39-60). Cambridge, MA: MIT Press.

Donagan, A. (1977). *The theory of morality*. Chicago: University of Chicago Press.

Enoch, D. (2006). Agency, shmagency: Why normativity won't come from what is constitutive of action. *Philosophical Review, 115*(2), 169-198.

Enoch, D. (2007). An outline of an argument for robust metanormative realism. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics: Volume 2*. Oxford: Oxford University Press.

Falk, W. D. (1948). Ought and motivation. *Proceedings of the Aristotelian Society, 48*, 111-138.

Foot, P. (1972). Morality as a system of hypothetical imperatives. *Philosophical Review, 81*, 305-316.

Frankena, W. K. (1958). Obligation and motivation in recent moral philosophy. In A. I. Melden (Ed.), *Essays in moral philosophy*. Seattle: University of Washington Press.

Frankena, W. K. (1976). In Goodpaster K. E. (Ed.), *Perspectives on morality: Essays of william K. frankena*. Notre Dame: University of Notre Dame Press.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy, 68*, 5-20.

Frankfurt, H. G. (1978). The problem of action. *American Philosophical Quarterly, 15*, 157-162.

Frankfurt, H. G. (1982). The importance of what we care about. *Synthese: An International Journal for Epistemology, Methodology and Philosophy of Science, 53*, 257-272.

Frankfurt, H. G. (1988). *The importance of what we care about: Philosophical essays*. New York: Cambridge Univesity Press.

Frankfurt, H. G. (1989). Concerning the freedom and limits of the will. *Philosophical Topics, 17*(1), 119-130.

Frankfurt, H. G. (2000). *Rationalism in ethics*. Berlin: Akademie Verlag.

Gaut, B. (1997). The structure of practical reason. In G. Cullity, & B. Gaut (Eds.), *Ethics and practical reason*. Oxford: Oxford University Press.

Gauthier, D. (1986). *Morals by agreement*. NY: Clarendon Press.

Gewirth, A. (1978). *Reason and morality*. Chicago: University of Chicago Press.

Gibbard, A. (1990). *Wise choices, apt feelings: A theory of normative judgment*. Cambridge: Harvard University Press.

Gibbard, A. (1999). Morality as consistency in living: Korsgaard's kantian lectures. *Ethics: An International Journal of Social, Political, and Legal Philosophy, 110*(1), 140-164.

Hammer, C. (2008). Explication, explanation and history. *History and Theory: Studies in the Philosophy of History, 47*(2), 183-199.

Hampton, J. (1998). *The authority of reason*. Cambridge: Cambridge University Press.

Harman, G. (1977). *The nature of morality: An introduction to ethics*. Oxford: Oxford University Press.

Harvey, M. T. (1998). Hobbes's deontological science of morals. (Ph.D., City University of New York).

Hobbes, T. (1651, 1968). In Macpherson C. B. (Ed.), *Leviathan.* Penguin.

Hume, D. (1739,1978). In Selby-Bigge L. A., Nidditch P. H. (Eds.), *A treatise of human nature* Oxford.

Joyce, R. (2002). *The myth of morality*. Cambridge: Cambridge University Press.

Kant, I. (1785,1911). *Grundlegung zur metaphysik der sitten*. Berlin: Georg Reimer Verlag.

Kant, I. (1788,1913). *Kritik der praktischen vernunft*. Berlin: Georg Reimer Verlag.

Korsgaard, C. (1985). Kant's formula of universal law. *Pacific Philosophical Quarterly, 66*, 24-47.

Korsgaard, C. (1986). Kant's formula of humanity. *Kant-Studien: Philosophische Zeitschrift Der Kant-Gesellschaft, 77*, 183-202.

Korsgaard, C. (1986). Skepticism about practical reason. *Journal of Philosophy, 83*, 5-25.

Korsgaard, C. (1989). Kant's analysis of obligation: The argument of foundations I. *Monist: An International Quarterly Journal of General Philosophical Inquiry, 72*, 311-340.

Korsgaard, C. (1997). The normativity of instrumental reason. In G. Cullity, & B. Gaut (Eds.), *Ethics and practical reason*. Oxford: Oxford University Press.

Korsgaard, C. (1999). Self-constitution in the ethics of plato and kant. *Journal of Ethics: An International Philosophical Review, 3*(1), 1-29.

Korsgaard, C. (2002). *Locke lectures.* Unpublished manuscript.from http://www.people.fas.harvard.edu/~korsgaar/

Korsgaard, C. (2007). Autonomy and the second person within: A commentary on stephen darwall's 'the second-person standpoint'. *Ethics: An International Journal of Social, Political, and Legal Philosophy, 118*(1), 8-23.

Korsgaard, C. M. (2003). Realism and constructivism in twentieth-century moral philosophy. *Philosophy in america at the turn of the century (APA centennial supplement journal of philosophical research)* (). Charlottesville: Philosophy Documentation Center.

Korsgaard, C. M. (2004). The myth of egoism. *Practical conflicts: New philosophical essays*. Cambridge: Cambridge University Press.

Korsgaard, C., Cohen, G. A., Geuss, R., Nagel, T., & Williams, B. (1996). In O'Neill O. (Ed.), *The sources of normativity*. Cambridge: Cambridge University Press.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong.* Penguin.

Mill, J. S. (1861,1979). In Sher G. (Ed.), *Utilitarianism.* Indianapolis: Hackett.

Moore, G. E. (1903). *Principia ethica*. Cambridge: Cambridge University Press.

Nagel, T. (1970). *The possibility of altruism*. Oxford: Oxford University Press.

Politis, V. (1997). Christine korsgaard's creating the kingdom of ends and the sources of normativity. *International Journal of Philosophical Studies, 5*(3), 425-448.

Price, R. (1757, 1969). A review of the principal questions in morals. In D. D. Raphael (Ed.), *British moralists 1650-1800* (pp. 655-762). Clarendon Press: Oxford University Press.

Prichard, H. A. (1928, 2002). Duty and interest. In J. MacAdam (Ed.), *Moral writings*. Oxford University Press.

Railton, P. (1986). Moral realism. *Philosophical Review, 95*, 163-207.

Railton, P. (1997). On the hypothetical and non-hypothetical in reasoning about belief and action. In G. Cullity, & B. Gaut (Eds.), *Ethics and practical reason*. Oxford: Oxford University Press.

Raphael, D. D. (Ed.). (1969). *British moralists 1650-1800, 2 volumes*. London: Oxford University Press.

Rawls, J. (1951). Outline of a decision procedure for ethics. *Philosophical Review, 60*, 177-197.

Rawls, J. (1971). *A theory of justice*. Cambridge MA: Belknap Press.

Rhodes, R. (1990). The moral leviathan. (Ph.D., City University of New York).

Rosati, C. (2003). Agency and the open question argument. *Ethics: An International Journal of Social, Political, and Legal Philosophy, 113*(3), 490-527.

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 49*, 329-359.

Scanlon, T. M. (1998). *What we owe to each other*. Cambridge, Mass: Belknap Press of Harvard University Press.

Schneewind, J. B. (Ed.). (1996). *Reason, ethics, and society*. Chicago: Open Court.

Schneewind, J. B. (1997). The sources of normativity. *Mind: A Quarterly Review of Philosophy, 106*(424), 106.

Setiya, K. (2003). Explaining action. *Philosophical Review, 112*(3), 339-393.

Shafer-Landau, R. (2006). Ethics as philosophy: A defense of ethical nonnaturalism. In T. Horgan, & M. Timmons (Eds.), *Metaethics after moore*. Oxford: Oxford University Press.

Shafer-Landau, R. (2003). *Moral realism: A defense*. Oxford: Oxford University Press.

Shafer-Landau, R. (2006). *Oxford studies in metaethics. vol. 1*. Oxford; New York: Oxford University Press.

Shafer-Landau, R. (2007). *Oxford studies in metaethics. vol. 2*. Oxford; New York: Oxford University Press.

Shafer-Landau, R. (2008). *Oxford studies in metaethics. vol. 3*. Oxford; New York: Oxford University Press.

Shafer-Landau, R., & Cuneo, T. (Eds.). (2007). *Foundations of ethics: An anthology*. Malden MA: Blackwell Publishing.

Smith, M. (1994). *The moral problem*. Cambridge: Blackwell.

Smith, M. (1995). Internal reasons. *Philosophy and Phenomenological Research, 55*(1), 109-131.

Smith, M. (1997). In defense of the moral problem: A reply to brink, copp, and sayre-McCord. *Ethics: An International Journal of Social, Political, and Legal Philosophy, 108*(1), 84-119.

Smith, M. (1998). The possibility of philosophy of action. In J. Bransen (Ed.), *Human action, deliberation and causation* (pp. 17-41). Dordrecht: Kluwer.

Smith, M. (1999). Search for the source (christine korsgaard's the sources of normativity). *Philosophical Quarterly, 49*(196), 384-394.

Smith, M. (2004). *Ethics and the A priori: Selected essays on moral psychology and meta-ethics*

Strawson, P. F. (1974). Freedom and resentment. *Freedom and resentment and other essays* (pp. 1-25). London: Methuen and co.

Street, S. (2008). Constructivism about reasons. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics: Volume 3* (). Oxford: Oxford University Press.

Sturgeon, N. L. (1984). Moral explanations. In D. Copp, & D. Zimmerman (Eds.), *Morality, reason and truth* (pp. 49-78). Totowa, N.J.: Rowman and Allanheld.

Taylor, A. E. (1938). The ethical doctrine of hobbes. *Philosophy, XIII*, 406-424.

Thalos, M. (2007). The sources of behavior: Toward a naturalistic, control account of agency. In D. Ross, D. Spurrett, H. Kincaid & G. L. Stephens (Eds.), *Distributed cognition and the will: Individual volition and social context* (pp. 123-168). Cambridge, MA: MIT Press.

Tiberius, V. (2008). *The reflective life : Living wisely with our limits*. Oxford; New York: Oxford University Press.

Velleman, J. D. (1985). Practical reflection. *Philosophical Review, 94*, 33-61.

Velleman, J. D. (1989). *Practical reflection*. Princeton: Princeton University Press.

Velleman, J. D. (1992). What happens when someone acts? *Mind: A Quarterly Review of Philosophy, 101*(403), 461-481.

Velleman, J. D. (1996). The possibility of practical reason. *Ethics: An International Journal of Social, Political, and Legal Philosophy, 106*(4), 694-726.

Velleman, J. D. (1997). Deciding how to decide. In G. Cullity, & B. Gaut (Eds.), *Ethics and practical reason*. Oxford: Oxford University Press.

Velleman, J. D. (2000). On the aim of belief. *The possibility of practical reason*. Oxford: Oxford University Press.

Velleman, J. D. (2000). *The possibility of practical reason*. New York: Oxford University Press.

Velleman, J. D. (2003). Narrative explanation. *Philosophical Review, 112*(1), 1-25.

Velleman, J. D. (2004). Précis of the possibility of practical reason. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 121*(3), 225-238.

Velleman, J. D. (2004). Replies to discussion on the possibility of practical reason. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 121*(3), 277-298.

Velleman, J. D. (2006). *Self to self: Selected essays*. Cambridge: Cambridge University Press.

Vogler, C. (2001). Anscombe on practical inference. In E. Millgram (Ed.), *Varieties of practical reasoning* (pp. 437-464). Cambridge: MIT Press.

Vogler, C. (2002). *Reasonably vicious*. Cambridge: Harvard University Press.

Wallace, R. J. (1994). *Responsibility and the moral sentiments*. Cambridge: Harvard University Press.

Wallace, R. J. (1997). Reason and responsibility. In G. Cullity, & B. Gaut (Eds.), *Ethics and practical reason*. Oxford: Oxford University Press.

Wallace, R. J. (2006). *Normativity and the will: Selected essays on moral psychology and practical reason*. Clarendon Press: Oxford.

Wallace, R. J. (2007). Reasons, relations, and commands: Reflections on darwall. *Ethics: An International Journal of Social, Political, and Legal Philosophy, 118*(1), 24-36.

Warrender, H. (1962). Hobbes's conception of morality. *Revista Critica Di Storia Della Filosofia, XVII*(IV), 434-449.

Watson, G. (1975). Free agency. *Journal of Philosophy, 72*, 205-220.

Watson, G. (2007). Morality as equal accountability: Comments on stephen darwall's 'the second-person standpoint'. *Ethics: An International Journal of Social, Political, and Legal Philosophy, 118*(1), 37-51.

Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge MA; Cambridge MA: Bradford Book; MIT Pr.

Wegner, D. M. (2004). Précis of the illusion of conscious will. *Behavioral & Brain Sciences, 27*(5), 649-659.

Williams, B. (1981). Internal and external reasons. *Moral luck* (pp. 101-113). Cambridge: Cambridge University Press.

Wilson, C. (2004). *Moral animals: Ideals and constraints in moral theory*. Clarendon Press: Oxford University Press.